Catalogue no. 89-653-X — No. 002
ISBN 978-1-100-22738-2

Concepts and Methods

Aboriginal Peoples Survey, 2012: Concepts and Methods Guide

by Elisabeth Cloutier and Éric Langlet

February 2014

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca,

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

To access this product

This product, Catalogue no. 89-653-X, is available free in electronic format. To obtain a single issue, visit our website, www.statcan.gc.ca, and browse by "Key resource" > "Publications."

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2014

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement (http://www.statcan.gc.ca/reference/copyright-droit-auteur-eng.htm).

Cette publication est aussi disponible en français.

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Table of contents

1. Introduction 7

1.1 Survey overview 7

1.2 Purpose of the Concepts and Methods Guide 7

2. Survey content: concepts and questions 8

2.1 Content development process 8

2.2 Aboriginal identity groups 9

A definition of Aboriginal identity 9

Questionnaire items for Aboriginal identification 9

Aboriginal identity variables available to data users 10

2.3 Levels of geography 11

2.4 Survey content themes 12

Education component 12

Employment component 13

Health supplement 13

Other core indicators 13

2.5 Questionnaire modules 13

2.6 Linked content from the National Household Survey 14

2.7 Content input from other surveys 15

Aboriginal identification 16

Education 16

Employment 16

Health 16

3. Survey design 16

3.1 Target population and coverage 16

3.1.1 Identifying the Aboriginal population 17

3.1.2 Survey reference date 19

3.1.3 National Household Survey frame 19

3.2 Sampling design 19

3.2.1 Domains of estimation 19

3.2.2 Stratification 22

3.2.3 Sampling design and sample allocation 23

3.2.4 Sample size adjustment 24

3.3 Sample size and response rate 25

Table of contents (continued)

4. Data collection 26

4.1 Time frame 26
4.2 Mode of collection 27
4.3 Supervision and quality control 27
4.4 Proxy interviews 27
4.5 Communications strategy 27
4.6 Special issues in the field 28
4.7 Final response rates 28

5. Data processing 28

5.1 Data capture 28
5.2 Social survey processing steps 28
5.3 Receipt of raw data and record clean up 29
5.3.1 Definition of a respondent 29
5.4 Variable recodes and multiple response questions 29
5.5 Flows: response paths, valid skips and question non-response 31
5.6 Coding 31
5.6.1 "Other-specify" items 31
5.6.2 Open-ended questions and standard classifications 32
5.7 Edit and imputation 32
5.8 Derived variables and NHS linkage 33
5.9 Creation of final data files and codebook (data dictionary) 33

6. Weighting 34

6.1 Initial weights 34
6.2 Adjustment for units not sent to collection 34
6.3 Adjustment for non-response 35
6.4 Adjustment for partial respondents 36
6.5 Post-stratification 36
6.6 Adjustment for extreme weights– Sigma gap method 37

Table of contents (continued)

7. Data quality 37

7.1 Overview of data quality evaluation 37
7.2 Sampling errors and bootstrap method 38
7.3 Non-sampling errors 40
7.3.1 Non-response errors 40
7.3.2 Coverage errors 41
7.3.3 Measurement errors 41
7.3.4 Processing errors 42

8. Differences between the Aboriginal Peoples Survey and other data sources 42

8.1 Differences between the Aboriginal Peoples Survey and the National Household Survey 42
8.1.1 Different interview methods and impact of proxy reporting 43
8.1.2 Different questionnaires 43
8.1.3 Different contexts 43
8.1.4 Effect of time 44
8.1.5 Processing 44
8.1.6 Differences in the estimates 44
8.2 Differences between the 2006 APS and the 2012 APS 46
8.2.1 Methodological differences 46
8.2.2 Content differences 46
8.2.3 Changes to Aboriginal identity questions 46

9. Data dissemination 48

9.1 An overview of 2012 APS dissemination 48
9.2 Data products and services 48
9.3 Analytical products 49
9.4 Reference products 49
9.5 Disclosure control 50

Appendix A 2012 Aboriginal Peoples Survey indicators by questionnaire module 51

Introduction 51
Aboriginal identification 51
Household composition and marital status 51
Mobility 51
Education 52
Aboriginal languages 54
Residential school 54

Table of contents (concluded)

Employment (aged 15 and over) 54
Traditional activities 55
Income 56
Health 56
Housing 57
Geography 57

Appendix B Extra response categories created for "Other – specify" questions 58

Appendix C Standard classifications used for the 2012 APS 59

Major field of study: Classification of Instructional Programs (CIP) 59
Industry and occupation: NAICS and NOCS 59
Aboriginal languages 59

Appendix D Glossary of survey terms 60

References 69

1. Introduction

1.1 Survey overview

The Aboriginal Peoples Survey (APS) is a national survey on the social and economic conditions of First Nations people living off reserve, Métis and Inuit aged 6 years and over. The objectives of the APS are to identify the needs of these Aboriginal groups and to inform policy and programs aimed at improving the well-being of Aboriginal peoples. The APS aims to provide current and relevant data for a variety of stakeholders, including First Nations, Métis and Inuit organizations, communities, service providers, researchers, governments and the general public.

The APS has been conducted by Statistics Canada since 1991, providing a range of social and economic indicators about Aboriginal peoples. It is a postcensal survey, designed to follow and complement the Census of Population and the National Household Survey (NHS). The 2012 APS represents the fourth cycle of the survey and the first to take a focused thematic approach. The focus for 2012 is on issues of education, employment and health. The survey will continue to provide core indicators in the areas of language, income, housing and mobility. Funding was provided by three federal departments: Aboriginal Affairs and Northern Development Canada, Health Canada and Employment and Social Development Canada (formerly called Human Resources and Skills Development Canada).

This cycle of the APS was conducted from February 6, 2012 to July 30, 2012. Over 50,000 people were selected to participate in the survey and the final response rate was 76%. The survey design allowed for the production of reliable data for each of the provinces and territories (Atlantic provinces grouped), as well as for each of the four Inuit regions: Nunatsiavut (Northern coastal Labrador), Nunavik (Northern Quebec), the territory of Nunavut and the Inuvialuit region of the Northwest Territories. The survey also targeted four particular education groups: current school attendees in grades 1 to 6; current school attendees in grades 7 to 12; high school completers (including equivalency); and high school leavers. Separate analyses on these dimensions are possible for each Aboriginal group: First Nations people living off reserve, Métis and Inuit aged 6 years and over.

Historically, data from the APS have been widely used and have filled data gaps on a wide range of questions. The data have been used to inform decision making in program and policy planning and development, to improve services for Aboriginal peoples and to support research. The Royal Commission on Aboriginal Peoples (RCAP) used the 1991 APS data as a primary source of demographic, social and economic information for their final report and related research studies. The Commission's final report recommended that the APS be conducted regularly to monitor the demographic and social conditions of Aboriginal peoples.

The 2012 APS will continue to inform policy and programming activities aimed at improving the well-being of Aboriginal peoples. The survey's new thematic data will help to provide a deeper understanding of the opportunities and challenges leading to success in education and employment. Information on health, language, income, housing and mobility will also be available for analysis. The APS will continue to serve as an important source of information for First Nations, Métis and Inuit organizations, community planners, service providers, governments and researchers.

1.2 Purpose of the Concepts and Methods Guide

This Concepts and Methods Guide is intended to provide a detailed review of the 2012 APS with respect to its subject matter and methodological approaches. It is designed to assist APS data users by serving as a guide to the concepts and measures of the survey as well as the technical details of the survey's design, field work and data processing. This guide is meant to provide users with helpful information on how to use and interpret survey results. The discussion on data quality also allows users to review the strengths and limitations of the data for their particular needs.

Chapter 1 of this guide provides an overview of the 2012 APS by introducing the survey's background and objectives. Chapter 2 outlines the survey's themes and explains the key concepts and definitions used for the survey. Chapters 3 to 6 cover important aspects of the APS survey methodology, from sampling design through data collection and processing. Chapters 7 and 8 review issues of data quality and caution users

about comparing 2012 APS data with data from other sources. Chapter 9 outlines the survey products available to the public, including data tables, analytical articles and reference material. Appendices provide a comprehensive list of survey indicators, extra coding categories and standard classifications used on the APS. A glossary of survey terms is also provided.

2. Survey content: concepts and questions

2.1 Content development process

The content for the 2012 Aboriginal Peoples Survey (APS) was developed by Statistics Canada in collaboration with the three federal funding departments: Aboriginal Affairs and Northern Development Canada, Health Canada and Employment and Social Development Canada (formerly called Human Resources and Skills Development Canada). An important framework guiding content development was the 2007 initiative on "Redefining how Success is Measured in Aboriginal Learning", conducted by the Canadian Council on Learning (CCL) in partnership with First Nations, Inuit and Métis communities and organizations across Canada. The 2012 APS also drew on many key indicators from previous cycles of the APS which were developed in collaboration with national Aboriginal organizations.

Following the release of data from the 2006 APS, an in-depth content review was conducted to ensure the future relevance of existing APS questions to key stakeholders and to identify potential data gaps. The review brought together expertise from a diverse group of researchers and subject matter experts from within and outside of Statistics Canada. This process included extensive analyses of the strengths and limitations of 2006 indicators and where refinements would be required. Indicators were reviewed in relation to 2006 data collection activities, coding, data processing steps and analytical results.

With respect to new content requirements for 2012, relevant standardized and well-established measures used on other Statistics Canada surveys were gathered and reviewed as potential indicators. These indicators increase opportunities to compare responses between the APS and other Statistics Canada surveys. In addition, recognizing the new CCL framework for Aboriginal learning, new APS content was sought that would allow for the measurement of a diverse range of educational experiences among Inuit, Métis and First Nations people living off reserve. The new CCL framework recognizes that Aboriginal learning is holistic and lifelong, comes from many different sources, and is rooted in Aboriginal languages and cultures (CCL, 2009).

These content development activities led to a series of recommendations that were implemented for the 2012 APS survey. The 2012 survey questionnaire then went through iterations of qualitative testing among First Nations people, Métis and Inuit across Canada, with a particular view to assessing cultural understanding of the questions. Qualitative testing was carried out by Statistics Canada's Questionnaire Design Resource Centre (QDRC) at eight locations across the country, including the North. Based on these results, adjustments were made to question wording, particularly for any new content, and to the flow of questions.

Prior to 2012, the APS used a paper questionnaire format. The questions in the 2012 APS were designed for use in a computer assisted interviewing (CAI) environment for the first time. CAI incorporated many features that served to maximize the quality of data collection. It enabled more complex question flows to be built in as well as on-line edits which identified logical inconsistencies that could be corrected by interviewers in collaboration with respondents at the time of the interview. Two computer-assisted interview questionnaires were developed for this survey: a Computer Assisted Telephone Interview (CATI) and a Computer Assisted Personal Interview (CAPI). These new CATI and CAPI questionnaires underwent extensive modular and end-to-end testing.

One final step was taken with respect to content development for the 2012 APS. Since the 2012 APS drew its sample from the 2011 National Household Survey (NHS) (see section 3.1 for details), it was decided that relevant information from the NHS would be combined with information provided during the 2012 APS interview through a process of data linkage. This approach would reduce the number of questions on the APS and provide for a more comprehensive dataset for APS users. More details of this process are provided in section 2.6.

2.2 Aboriginal identity groups

First Nations people living off reserve, Métis and Inuit constitute what is called the "target population" for the 2012 Aboriginal Peoples Survey. The survey data will support distinct analyses for each of these groups.

The 2012 APS did not include First Nations people who were living in First Nations communities (on reserve) at the time of data collection. For current information on First Nations people living on reserve, please refer to the 2011 National Household Survey and to a non-Statistics Canada survey: the First Nations Regional Early Childhood, Education and Employment Survey (FNREEES). The FNREEES will be conducted by the First Nations Information Governance Centre starting in the fall of 2013. More information on the FNREEES is available from the First Nations Information Governance Centre website.

For the purposes of the 2012 APS, First Nations people living off reserve, Métis and Inuit were defined on the basis of an "Aboriginal identity" concept. In both conceptual and methodological terms, Aboriginal identity has many complexities. This section of the guide addresses the definition of Aboriginal identity and how it was measured on the 2012 APS questionnaire. A more technical description of the survey sampling methods used to identify the Aboriginal identity population for the 2012 APS is provided in chapter 3. That chapter includes details on how APS respondents were chosen from respondents to the 2011 NHS.

A definition of Aboriginal identity

The 2012 Aboriginal Peoples Survey identifies the Aboriginal identity population as anyone who reported being at least one of the following:

- an Aboriginal person, that is, First Nations (North American Indian), Métis or Inuk (Inuit);
- a Status Indian, that is, a Registered or Treaty Indian as defined by the *Indian Act* of Canada;
- a member of a First Nation or Indian band

The first criterion above is referred to as "Aboriginal self-reporting". A respondent may self-report as being an Aboriginal person and/or they may see themselves as belonging to one or more of the particular Aboriginal groups mentioned: First Nations, Métis or Inuit. First Nations includes Status and Non-Status Indians. It should also be noted that some respondents use the term First Nations while others use the term North American Indian.

As the APS definition implies, a person need not self-report as Aboriginal or as First Nations, Métis or Inuk in order to be considered part of the Aboriginal identity population. If a person has reported being a Status Indian or a member of a First Nation or Indian band, they are considered to be part of the Aboriginal identity population, regardless of their responses to other questions.

With respect to the measurement of "Status Indian", the 2012 APS includes everyone who said that they are a Registered or Treaty Indian as defined by the *Indian Act*. This definition includes those who were registered under Bill C-31 and Bill C-3, amendments made to the *Indian Act* in 1985 and 2011, respectively. The newest 2011 amendment, Bill C-3, is called the *Gender Equity in Indian Registration Act*.

Questionnaire items for Aboriginal identification

Table 1 below lists the six Aboriginal identification questions asked in the 2012 APS and shows the Aboriginal identity classification derived from the answers provided to these questions. As shown, in order for someone to be part of the Aboriginal identity population, respondents needed to have a YES response to ID_Q01 (self-reported Aboriginal person) or a YES response to at least one of questions ID_Q03 (Status Indian), ID_Q04B (Registered as a Status Indian under Bill C-31 or Bill C-3), or ID_Q05 (Member of First Nation or an Indian band). Note that those who answered NO to all these questions were classified as non-Aboriginal and considered out-of-scope for the survey. Adjustments to survey weight were made accordingly.

As shown in table 1, the measurement of Aboriginal self-reporting was divided into two parts: ID_Q01 and ID_Q02. This allowed respondents to self-report as an Aboriginal person in question ID_Q01 even if they did not self-report as a specific group as named in ID_Q02 (First Nations, Métis or Inuit). For the 2012 APS, respondents who said YES to ID_Q01 but NO to ID_Q02 were classified to a specific group based on their responses to the 2011 National Household Survey. In addition, those who reported that they were a Status Indian or a member of a First Nation or Indian band without a self-reported Aboriginal group were classified as having a First Nations identity for the purposes of the 2012 APS.

For more information with respect to the collection and processing of Aboriginal identification data in the 2012 APS, please also refer to sections 5.7 of this guide. As well, for differences between the 2012 APS and 2006 APS questions on Aboriginal identity, see section 8.2.3.

Table 1 Questionnaire flow and Aboriginal identity classification

Questionnaire flow and Aboriginal identity classification	First Nations	Métis	Inuit	Non-Aboriginal (out of scope)
ID_Q01 - Are you an Aboriginal person, that is, First Nations, Métis or Inuk (Inuit)? First Nations includes Status and Non-Status Indians.				
Yes ... see subsequent questions				
ID_Q02 - Are you First Nations, Métis or Inuk (Inuit)?				
First Nations				
(North American Indian)	X			
Métis		X		
Inuk			X	
No - see subsequent questions				
ID_Q03 - Are you a Status Indian, that is, a Registered or Treaty Indian as defined by the Indian Act of Canada?				
Yes	X			
No - classification based on overall responses				
ID_Q04A - Have you ever applied to Aboriginal Affairs and Northern Development Canada (previously named INAC) to be registered as a Status Indian under Bill C-31 or Bill C-3?				
Yes ...				
ID_Q04B - Have you been registered as a Status Indian under Bill C-31 or Bill C-3?				
Yes, Bill C-31	X			
Yes, Bill C-3	X			
No - classification based on overall responses				
No - classification based on overall responses				
ID_Q05 - Are you a member of a First Nation or Indian band?				
Yes	X			
No - classification based on overall responses				
No to all questions: ID_Q01-Q02, ID_Q03, ID_Q04A-Q04B and ID_Q05				X

Aboriginal identity variables available to data users

Data made available from the 2012 APS will provide analytical variables for each aspect of Aboriginal identity. These variables will be central to data users for conducting their analyses of subject matter themes for each group - First Nations people living off reserve, Métis and Inuit. Aboriginal identity variables will include indicators of both single and multiple identities (for example, persons who reported as being both a First Nations person and Métis). Variables on Status Indian (Registered or Treaty) will also be available for analysis. A variable on Aboriginal ancestry will be provided from the NHS since it was not measured directly by the 2012 APS.

2.3 Levels of geography

The 2012 Aboriginal Peoples Survey ensured coverage of certain core geographic domains. These included provinces and territories (with the Atlantic provinces grouped) and the four Inuit regions of Inuit Nunangat.

Inuit Nunangat is the homeland of Inuit of Canada. It includes the communities located in the four Inuit regions: Nunatsiavut (Northern coastal Labrador), Nunavik (Northern Quebec), the territory of Nunavut and the Inuvialuit region of the Northwest Territories. These regions collectively encompass the area traditionally occupied by Inuit in Canada (see map 1).

As described in more detail in section 3.2 covering sampling design, these geographic domains were targeted by the 2012 APS to ensure that adequate data estimates would be available at these levels of geography. Other geographic variables are also available in the 2012 APS database, based on geographies from the National Household Survey. These include Census Metropolitan Areas (CMAs), Census Subdivisions (CSDs) and Population Centres, among others. In addition, geographies will include health regions across Canada which represent administrative areas or regions of interest to health authorities.

Users should note, however, that not all APS data can be cross-tabulated or analyzed at detailed levels of geography. Some data tables will be possible for more detailed geographies but the reliability of data estimates at each level will need to be examined on a case by case basis.

Data users are directed to the 2012 APS codebook (data dictionary) for a complete list of geographies available from the Aboriginal Peoples Survey (contact Statistics Canada Client Services at sasd-dssea@statcan.gc.ca or 1-800-263-1136). The *National Household Survey (NHS) Dictionary* also defines geographies relevant to the APS.

Map 1 The four regions of Inuit Nunangat

Source(s): Inuit Tapiriit Kanatami

2.4 Survey content themes

Reflecting a focused thematic approach, the 2012 Aboriginal Peoples Survey collected data falling into four content areas:

- Education Component
- Employment Component
- Health Supplement
- Additional Core Content

This section describes these four content areas. Section 2.5 below provides an outline of the questionnaire modules of the 2012 APS. Appendix A provides a detailed list of the indicators measured in each of these modules. For more information, refer to the 2012 APS questionnaire.

In addition, a comprehensive description of all the variables available from the survey data is provided in the 2012 APS codebook (data dictionary). For details on how to obtain the codebook, please contact Statistics Canada Client Services at sasd-dssea@statcan.gc.ca or 1-800-263-1136.

Education component

The content for the education component of the 2012 APS was conceptualized based on different educational groups or streams. Questions were shaped to inquire into the school experiences of persons in four different educational groups:

- Current school attendees in grades 1 to 6;
- Current school attendees in grades 7 to 12;
- High school completers, with a diploma or its equivalent; and
- High school leavers with no high school diploma and not currently attending grades 1-12

High school completers are defined as persons who had met the minimum requirements for high school graduation at the time of the survey. This group includes persons who had completed a high school equivalency program. Leavers are defined as those who were not in elementary or high school at the time of the survey and had not met the minimum requirements for high school graduation. Persons enrolled in a high school equivalency program but had not completed it at the time of the survey were grouped as "leavers".

The survey followed different questionnaire flows for the different educational groups, providing data on selected themes for each group, as follows:

- For current attendees: school experiences of children and youth, school characteristics, parental involvement, educational aspirations
- For high school completers and leavers: elementary and high school experiences, particularly in last year of school; educational pathways through the school system, highest level of schooling attained, barriers encountered, post-secondary school experiences

Content for each of these groups has led to a rich data set for addressing a range of potential research and policy questions as well as educational program needs. The survey provides data on the current needs of Aboriginal students and their families and also allows for a broader understanding of the factors associated with the successful completion of high school. The data allow for in-depth analysis of the differences between high school completers and leavers in terms of the quality of school experiences and the difficulties students face, including the types of support students receive at school, family involvement, relationships with peers, the experience of changing schools, and reasons for leaving school. Educational data also cover other important indicators such as highest level of schooling, post secondary credentials and major field of study.

It should also be noted that even though the APS targets First Nations people living off reserve, some of these respondents may have gone to school on a reserve at some point in the past. The 2012 APS asks respondents if they have ever attended an elementary or high school in a First Nations Community (on reserve).

Employment component

The employment component of the 2012 APS focuses on persons aged 15 and over. It explores employment and unemployment experiences, covering key labour force indicators such as labour market activities in a specific reference period, occupation and industry of work, and non-standard work arrangements such as temporary work, part-time work and the holding of multiple jobs. The content also covers in-depth data on potential labour market difficulties including barriers encountered in finding work, reasons for working part-time, methods used in looking for work, and impediments to labour mobility. Participation in traditional activities is also captured.

This content provides rich data for addressing a range of research and policy issues in the area of employment for First Nations people living off reserve, Métis and Inuit. These indicators also enrich analyses of successes and challenges in relation to high school completers and leavers.

Health supplement

The 2012 APS Health Supplement covers the important topics of access to health care, measures of physical health, mental health, injuries, smoking, drinking, and drug use, food insecurity and community support. These indicators provide for potential analyses of issues related to health care access, physical and mental health status and areas of health risks. These indicators will also enrich the analysis of successes and barriers in education and employment.

Other core indicators

A wide range of other core indicators have been added to the 2012 APS covering the areas of Aboriginal languages, housing, mobility, household composition and residential school attendance. These variables will allow for ongoing tracking of many important issues for First Nations people living off reserve, Métis and Inuit.

2.5 Questionnaire modules

This section provides a list of the modules on the 2012 APS questionnaire. Appendix A provides a detailed list of the indicators measured in each of these modules. Appendix B lists the extra classification categories created during survey coding. Appendix C describes the standard classifications used to create indicators for open-ended survey questions.

- Introduction
- Aboriginal identification
- Household composition and marital status
- Mobility
- Education
  - Education status
  - Education 1: Current attendees, grades 1 to 6
  - Education 2: Current attendees, grades 7 to 12
  - Education 3: High school leavers and completers, under age 45
  - Education 4: High school leavers and completers, age 45 and over
- Aboriginal languages
- Residential school
- Employment (aged 15 and over)
  - Labour market activities
  - Labour force status
  - Looking for work
  - Discouraged workers

  - Past job attachment
  - Multiple employment
  - Class of worker
  - Incorporated business
  - Job tenure
  - Industry
  - Occupation
  - Usual hours of work
  - Part-time employment
  - Permanent work
  - Labour mobility
- Traditional activities
- Income
- Health
  - General health 1 – Health status
  - Pregnancy and childbirth
  - Height and weight
  - General health 2 – Contact with health professionals
  - Chronic conditions
  - Injuries
  - Mental health
  - Distress
  - Suicide
  - Smoking
  - Alcohol use
  - Drug use
  - Food security
  - Community support
- Housing

2.6 Linked content from the National Household Survey

The Aboriginal Peoples Survey (APS) has historically been a postcensal survey, with its sample being drawn from Aboriginal census respondents. In 1991, 2001 and 2006, APS data were linked with data from each corresponding Census of Population. In 2011, the Census of Population was modified and the very detailed information previously collected on the long-form Census questionnaire was moved to the new 2011 National Household Survey (NHS), including the questions identifying the Aboriginal population. For 2012, the APS sample was drawn from respondents to the 2011 NHS and the final Aboriginal Peoples Survey microdata file was linked with the 2011 National Household Survey Dissemination Database. (Additional information on the sample selection for the APS is provided in Chapter 3 of this document).

The specific benefits of an APS-NHS record linkage are reduced response burden for the target population of the APS, the establishment of survey weights which are crucial to providing valid estimates, and the creation of a comprehensive microdata file which can be used by data analysts to extend their learning and to inform policy and program development for Aboriginal Peoples in Canada. Together, data from the two sources will provide a detailed statistical portrait of First Nations people living off-reserve, Métis and Inuit in Canada - data which are not available from any other source.

At the time of data collection, all NHS respondents were informed that the information they provided might be used in support of other Statistics Canada surveys. Specifically, the message on the cover of the 2011 NHS questionnaire stated: "Your information may be used by Statistics Canada in support of our other surveys or for analysis." As well, at the outset of the 2012 APS survey interview, respondents were told about the purpose of the survey, its voluntary nature and that:

"In order to reduce the number of questions, Statistics Canada plans to combine information collected during the 2011 National Household Survey to the information you provide in this interview. We may also add information from other surveys or administrative data sources. All information will be kept confidential and used for statistical purposes only".

All products containing linked data are disseminated in accordance with Statistics Canada's policies, guidelines and standards. Only aggregate statistical estimates that conform to the confidentially provisions of the *Statistics Act* can be released outside of Statistics Canada. All data requests for APS tabulations are screened for confidentiality and the aggregate data are rounded before being released to clients.

More than 100 NHS variables were linked to the final APS file for 2012. The list below indicates the type of NHS variables that have been appended to the APS analytical file. It is important to note that these NHS variables, having been obtained from the 2011 NHS responses for 2012 APS respondents, refer to each respondent's situation on the day of the NHS, that is, as of May 10, 2011. Users should be aware that in some cases, the respondent may have moved, had a change in the composition of their household, or had a change in employment between the date of the NHS and the date of the APS interview. In other words, some of the information provided by the NHS may not be reflective of the respondent's situation when the APS interview took place.

A complete list of linked NHS variables and their accompanying notes are provided in the 2012 APS codebook (data dictionary) which accompanies the APS analytical file.

- **Household level variables**
  - Geography, including census metropolitan areas and Inuit area of residence
  - Housing, including tenure, number of rooms in dwelling and need for repairs
  - Family, including presence of spouse/partner in household, presence and number of children in census family and parent or guardian information for child respondents
  - Family, including presence of spouse/partner in household, presence and number of children in census family, number of persons in the census family, and census family status and structure. In addition, a series of derived variables is included in order to provide NHS information on the parent(s) or guardian(s) of APS respondents who were less than 18 years of age. These variables include several NHS characteristics of the first person most knowledgeable about the respondent (PMK1) and, if applicable, the same characteristics for second person most knowledgeable (PMK2, who was generally the spouse or partner of PMK1). NHS characteristics of the PMK1 and PMK2 included Aboriginal identity and ancestry, Registered or Treaty Indian status, age and sex, marital and common-law status, Aboriginal mother tongue and Aboriginal home language, labour force status, and total income.
- **Person-level variables**
  - Ethnicity and religion
  - Employment, including labour force status, weeks worked in 2010, class of worker, full-time or part-time work
  - Place of work, including place of work status, type of commuting and distance
  - Mobility, including mobility status 1 and 5 years ago
  - Income, including family income, employment income, low-income status
  - Language, including knowledge of official languages, mother tongue, language spoken at home and language of work
  - Activity limitations, including activity difficulties/reductions at home, work and school

2.7 Content input from other surveys

The 2012 APS drew on questionnaire content from multiple sources. The 2006 APS was used as a starting point and served as a key source of well-established questions for use with First Nations people living off-reserve, Métis and Inuit. Questions were also drawn from other surveys as much as possible to enhance robustness of the data and allow for comparability across data sources, where appropriate. Sometimes these questions were modified for a better fit with the APS. In such instances, the data are not directly comparable.

This section outlines the primary sources used for 2012 APS questionnaire development in relation to the main survey themes. It should be noted that these sources do not represent a comprehensive list of all the Statistics Canada surveys that include indicators found on the APS.

Aboriginal identification

- 2011 National Household Survey (NHS)
- 2006 Aboriginal Peoples Survey (APS)

Education

- 2006 APS, 2001 APS, and 1991 APS
- 2006 Aboriginal Children's Survey (ACS)
- National Longitudinal Survey of Children and Youth (NLSCY)
- Youth in Transition Survey (YITS)
- Program for International Student Assessment (PISA)
- National Household Education Surveys Program
- Access and Support to Education and Training Survey (ASETS)
- School Leavers Survey

Employment

- 2006 APS
- Labour Force Survey (LFS)
- Survey of Older Workers (SOW)

Health

- 2006 APS
- First Nations Regional Health Survey (RHS)
- Canadian Community Health Survey (CCHS)
- Survey of Young Canadians (SYC)
- Canadian Health Measures Survey (CHMS)

It is also worth noting that there are four main social surveys at Statistics Canada for which data are available by Aboriginal group: the National Household Survey (NHS), the Labour Force Survey (LFS), the General Social Survey (GSS) and the Canadian Community Health Survey (CCHS).

3. Survey design

3.1 Target population and coverage

The target population of the 2012 Aboriginal Peoples Survey (APS) was composed of the Aboriginal identity population of Canada, 6 years of age and over as of February 1, 2012, living in private dwellings excluding people living on Indian reserves and settlements and in certain First Nations communities in Yukon and the Northwest Territories (NWT). These exclusions were made to avoid overlap with the First Nations Regional Early Childhood, Education and Employment Survey to be conducted in the fall of 2013 by the First Nations Information Governance Centre (see section 2.2). The census subdivisions (CSD) covered in Yukon and the NWT are listed in Table 2.

Table 2 List of census subdivisions in Yukon and Northwest Territories (NWT)

Territory	CSD No.	Name of CSD	CSD type
Yukon	6001003	Watson Lake	T
	6001004	Faro	T
	6001009	Whitehorse	CY
	6001029	Dawson	T
	6001044	Mt. Lorne	HAM
	6001045	Yukon, Unorganized	NO
	6001046	Swift River	SÉ
	6001049	Destruction Bay	SÉ
	6001050	Stewart Crossing	SÉ
	6001052	Keno Hill	SÉ
	6001055	Ibex Valley	HAM
	6001058	Marsh Lake	NO
	6001059	Macpherson-Grizzly Valley	NO
	6001060	Whitehorse, Unorganized	NO
NWT	6101014	Paulatuk	HAM
	6101017	Inuvik	T
	6101025	Aklavik	HAM
	6101036	Tuktoyaktuk	HAM
	6101041	Sachs Harbour	HAM
	6101063	Region 1, Unorganized	NO
	6101095	Ulukhaktok	HAM
	6102007	Norman Wells	T
	6102063	Region 2, Unorganized	NO
	6103097	Region 3, Unorganized	NO
	6104097	Region 4, Unorganized	NO
	6105003	Enterprise	SET
	6105016	Hay River	T
	6105026	Reliance	SET
	6105097	Region 5, Unorganized	NO
	6106023	Yellowknife	CY
	6106097	Region 6, Unorganized	NO

The CSD types can be found in Table 5 of the Census Dictionary. The "Aboriginal identity population" is defined in section 3.1.1, below.

3.1.1 Identifying the Aboriginal population

The APS is a survey that selects its sample from respondents who reported either Aboriginal identity or Aboriginal ancestry to the 2011 National Household Survey (NHS) questionnaire. Although the NHS replaced the 2006 Census long form, the APS is still considered a postcensal survey, given the link between the NHS and the Census. More precisely, the APS sample was selected from respondents who gave specific answers to four screening questions on the NHS questionnaire, which had two main versions, the N1 form and the N2 form.

The N1 form was completed by self-enumeration and was administered to approximately one in three households in most parts of Canada (N1 regions). Other than the basic census demographic questions (name, sex, date of birth, legal marital status, common-law status, relationship to person 1, various language questions and the consent question to release the data in 92 years), the NHS N1 form included questions on labour market activity, income, education, activity limitations, citizenship, housing, ethnic origin, and so on.

The N2 form, identical in content to the N1 form, except for some adapted examples and excluded questions[1], was administered by personal interview to all households in remote areas, Inuit communities and Indian reserves and settlements (N2 regions).

1. The questions on citizenship (question 10), immigrant status received (question 11) and year of immigration (question 12) are not asked of people living on Indian reserves and settlements who were enumerated using the N2 form.

The four screening questions used to identify the Aboriginal population were ethnic origin (question 17), Aboriginal self-reporting (question 18), Status Indian (Registered or Treaty Indian, question 20), and First Nation / Indian band membership (question 21).

The reporting of an Aboriginal origin in question 17 defines the Aboriginal ancestry population (or ancestry population).

The Aboriginal identity population of the NHS is derived from three questions: NHS questions 18, 20 and 21. The concept of Aboriginal identity refers to those persons who either (a) self-reported as at least one Aboriginal group, namely, First Nations (North American Indian), Métis or Inuit; and/or (b) reported being a Status Indian (Registered Indian or Treaty Indian, as defined by the *Indian Act of Canada*); and/or (c) reported being a member of a First Nation or Indian band.

Within the context of the APS, individuals with an Aboriginal ancestry who did not have an Aboriginal identity in the NHS are defined as the Aboriginal ancestry-only population (or ancestry-only population). For the purposes of APS sampling, the Aboriginal population includes both the identity population and the ancestry-only population. Although, in contrast to the 2006 APS, the ancestry-only population was not part of the 2012 APS target population, it was still sampled in the NHS because it was noted that in 2006, slightly less than one-third of the 2006 Census ancestry-only population had reported Aboriginal identity in the 2006 APS.

The Aboriginal identity concept in the APS is the same as that in the NHS, but is defined based on a slightly different set of questions (see Table 1 in section 2.2).

National Households Survey – N1 - question 17

What were the ethnic or cultural origins of this person's ancestors?

This question collects information on the ancestral origins of the population and provides information about the composition of Canada's diverse population.

An ancestor is usually more distant than a grandparent.

For example, Canadian, English, French, Chinese, East Indian, Italian, German, Scottish, Irish, Cree, Mi'kmaq, Salish, Métis, Inuit, Filipino, Dutch, Ukrainian, Polish, Portuguese, Greek, Korean, Vietnamese, Jamaican, Jewish, Lebanese, Salvadorean, Somali, Colombian, etc.

1: Specify as many origins as applicable using capital letters.

2011 National Household Survey – N1 - question 18

Is this person an Aboriginal person, that is, First Nations (North American Indian), Métis or Inuk (Inuit)?

Note: First Nations (North American Indian) includes Status and Non-Status Indians.

If " Yes ", mark the circle(s) that best describe(s) this person now and go to Question 20.

1: No, not an Aboriginal person

2: Yes, First Nations(North American Indian)

3: Yes, Métis

4: Yes, Inuk (Inuit)

2011 National Household Survey – N1 - question 20

Is this person a Status Indian (Registered or Treaty Indian as defined by the Indian Act of Canada)?

1: No

2: Yes, Status Indian (Registered or Treaty)

2011 National Household Survey – N1 - question 21

Is this person a member of a First Nation/Indian band?

If " Yes ", which First Nation/Indian band?

For example, Musqueam Indian Band, Sturgeon Lake First Nation, Atikamekw of Manawan.

1: No

2: Yes, member of a First Nation/Indian band (Specify name of First Nation/Indian band.)

3.1.2 Survey reference date

February 1, 2012 was used as the APS reference date. This date corresponds approximately to the beginning of data collection for the survey. Age is established based on this reference date and determines the questionnaire flow to be used.

3.1.3 National Household Survey frame

The APS sample was selected from the unedited, non-imputed database of the National Household Survey (NHS), which is the NHS database referred to as the Response DataBase or RDB. Before selecting the sample, survey methodologists developed an editing and imputation strategy to deal with missing values in any of the four NHS screening questions (or "filter questions") or in the variables used during stratification (including age and certain education variables), as well as for individuals with certain characteristics which appeared to be inconsistent with being Aboriginal.

The first step in selecting the sample was to include on the survey frame all individuals reporting Aboriginal identity or ancestry who were aged 6 and older as of February 1, 2012.

In the second step, all individuals who were part of the same households as the units in the initial frame were added to the frame. These additional persons were retained on the frame only if their responses to screening questions were missing but the characteristics of the people in the household who had answered the screening questions indicated that they would have had a good chance of having Aboriginal identity and/or ancestry. Hence, an individual with missing screening questions on identity would normally have been imputed as having Aboriginal identity if at least 50% of the members of the same household who completed the identity screening questions had Aboriginal identity. Similarly, an individual with non-response to the ancestry screening question would normally have been imputed as having Aboriginal ancestry if at least 50% of the members of the same household who completed the screening question on ancestry had Aboriginal ancestry.

It was clearly important to stratify the 2012 APS sample by the education groups described in section 2 (current attendees, grades 1 to 6; current attendees, grades 7 to 12; high school completers or leavers). Estimates were targeted for each of these four groups. Because a person's education status may have changed between the time of the 2011 NHS and February 1, 2012 a deterministic imputation model was developed to determine the most likely education group for each respondent as of February 1, 2012.

Once the processing was complete, individuals under 6 years of age as of February 1, 2012, those living on reserves or in certain specific communities in Yukon and the Northwest Territories and individuals who became non-Aboriginal following processing were excluded from the survey frame.

3.2 Sampling design

3.2.1 Domains of estimation

An effective stratification uses domains of estimation. Domains of estimation are groups of units for which estimates are targeted. In the case of the APS, *stratification-specific domains of estimation* were used. These domains of estimation corresponded to geographical regions for which estimates with an "acceptable" level of precision for a particular Aboriginal group (i.e. First Nations, Métis or Inuit) and particular education group were targeted.

An example of a domain of estimation would be Métis in Alberta attending elementary school. During stratification, the Métis Aboriginal group was comprised of individuals reporting Métis identity alone to NHS question 18 or, for individuals without Aboriginal identity (Aboriginal ancestry-only population), Métis ancestry alone to question 17 (with or without non-Aboriginal ancestry). In reality, Aboriginal ancestry-only individuals were not part of the survey's target population but were sampled because they had a relatively important chance of reporting identity on the survey as described in section 3.1.1. This is why the term "*stratification-specific domains of estimation*" is used rather than the term "*survey-specific domains of estimation*".

More precisely, the stratification-specific domains of estimation were created by cross-tabulating the following variables:

- Geography
    - Inuit regions
    - Outside Inuit regions
        - province/territory
        - Atlantic provinces grouped
- Education group
    - Current attendees, elementary school (grades 1 to 6)
    - Current attendees, high school (grades 7 to 12)
    - Completers: high school diploma or equivalent
    - Leavers: no high school diploma or equivalent and not currently attending elementary or high school
- Aboriginal group
    - Inuit in Inuit regions
    - Inuit outside Inuit regions (rest of Canada)
    - Aboriginal groups combined for Atlantic Canada, Quebec (outside Nunavik), Yukon and Northwest Territories (outside Inuvialuit)
    - For Ontario, Manitoba, Saskatchewan, Alberta and British Columbia
        - Status First Nations people living off-reserve
        - Non-Status First Nations people living off-reserve
        - Métis

Here is how the Aboriginal groups were defined during stratification based on responses to the NHS:

- Status First Nations people living off-reserve – persons answering First Nations alone to NHS question 18 and answering YES to NHS question 20
- Non-Status First Nations people living off-reserve – persons answering First Nations alone to question 18 and answering NO to question 20 or persons without Aboriginal identity but responding First Nations alone to question 17 (with or without non-Aboriginal ancestry)
- Métis only – persons answering Métis only to question 18 or persons without Aboriginal identity but answering Métis alone to question 17 (with or without non-Aboriginal ancestry)
- Inuit only – persons answering Inuit only to question 18 or persons without Aboriginal identity but answering Inuit alone to question 17 (with or without non-Aboriginal ancestry)
- Multiple Aboriginal group – persons reporting more than one identity to question 18 or persons without Aboriginal identity but reporting more than one Aboriginal ancestry to question 17 (with or without non-Aboriginal ancestry)
- Status Indian or member of a First Nation / Indian band only – persons answering YES to question 20 or to question 21 but NO to question 18.

To ensure the reliability of estimates, some education groups and/or Aboriginal groups had to be combined in certain regions. Education groups were combined to reduce the number from four to two by grouping those attending elementary school with those attending high school, and completers with leavers. In Atlantic Canada outside Nunatsiavut, in Quebec outside Nunavik, in Yukon and in the NWT, the number of Métis is generally too small to be able to produce separate estimates for First Nations people and Métis by education group. For this reason, all Aboriginal groups were combined during stratification.

Note that although estimates were targeted for Inuit outside Inuit regions nationally, domains of estimation were "artificially" created for Inuit in the group of provinces from Ontario to British Columbia combined (for the other provinces and territories outside Inuit regions, all Aboriginal groups were combined). Given the small group that Inuit represent between Ontario and British Columbia, relatively imprecise estimates were targeted for two education groups only. The estimates for Inuit outside Inuit regions nationally were expected to be more precise when Inuit from other provinces/territories outside Inuit regions were added.

In total, 88 target domains of estimation and 92 supplementary domains of estimation were created, for a total of 180 domains of estimation. Supplementary domains of estimation included domains where there was no plan to produce estimates for the APS but where a particular group should still be represented in the sample. Examples of supplementary domains of estimation included the non-Inuit population living in Inuit regions and the population aged 45 years and older.

For each target domain, the goal was to estimate a characteristic present for at least 10% of individuals in the education groups associated with children and youth (currently attending grades 1 to 6 and currently attending grades 7 to 12) and for at least 7.5% of individuals in the education groups associated with adults (completers and leavers), with a given coefficient of variation (CV). The coefficient of variation is a measure of the precision of the estimate, which is described in section 7.2 (*Sampling errors and bootstrap method*). The minimum proportion targeted is referred to as "*min-p*". The ability to achieve the targeted CVs for a given *min-p* value depended on such factors as the size of the population, the number of respondents available from the NHS, the expected response rate, the expected number of false positives (persons reporting Aboriginal identity or ancestry in the NHS without Aboriginal identity in the APS, a very specific type of respondents outside the target population), the sample loss associated with the constraint of selecting a maximum of three persons per household, and the loss associated with reducing overlap with other surveys. Depending on the domain sizes, CV values were set to 20%, 22.5%, 25% or 33%.

The following table gives the targeted CVs and *min-p* for each target domain of estimation.

Table 3 Coefficients of variation and *min-p* values by domain of estimation

	Aboriginal group					Education group					
Region	Status FN	Non-Status FN	Métis	Inuit	All	Grades 1 to 6 (*min-p*=10%)	Grades 7 to 12 (*min-p*=10%)	Completer (*min-p*=7.5%)	Leaver (*min-p*=7.5%)	Grades 1 to 12 (*min-p*=10%)	Completer or leaver (*min-p*=7.5%)
						percentage					
Nunatsiavut				x						20	20
Nunavik				x		20	20	20	20		
Inuvialuit				x						20	20
Nunavut				x		20	20	20	20		
Inuit Ont-BC				x						33	33
Atlantic outside Nunatsiavut					x	20	20	20	20		
Quebec outside Nunvik					x	20	20	20	20		
Ontario	x				x	25	25	25	25		
		x				25	25	25	25		
			x			22.50	22.50	22.50	22.50		
Manitoba	x					25	25	25	25		
		x				33	33	33	40		
			x			22.50	22.50	22.50	22.50		
Saskatchewan	x					25	25	25	25		
		x				33	33	33	40		
			x			22.50	22.50	22.50	22.50		
Alberta	x					25	25	25	25		
		x				25	25	25	33		
			x			22.50	22.50	22.50	22.50		
BC	x					25	25	25	25		
		x				25	25	25	33		
			x			22.50	22.50	22.50	22.50		
Yukon					x					25	25
NWT outside Inuvialuit					x	25	25	25	25		

For example, in Nunatsiavut, estimates were targeted for Inuit in only two education groups: current attendees in grades 1 to 12 and completers and leavers combined. The intention was to estimate, with a CV of 20%, a minimum proportion of 10% in the first group and 7.5% in the second group. In Ontario, estimates were targeted for Status and Non-Status First Nations people separately, as well as for Métis, across the four education groups. CVs of 25% were targeted for Status and Non-Status First Nations people and CVs of 22.5% are targeted for Métis, within each education group.

Clearly, any combination of domains will give smaller CVs. Because separate estimates were targeted for Status and Non-Status First Nations people for each province between Ontario and British Columbia, CVs of 25% were set for these two groups, which should produce CVs smaller than 20% for First Nations with or without Status. CVs of 22.5% were targeted for Métis in each province between Ontario and British Columbia.

3.2.2 Stratification

Stratification will produce more precise estimates if units are homogeneous within strata and heterogeneous between strata. In addition, the estimation weights associated with survey respondents should ideally be as close as possible within strata.

The NHS design is a two phase design in regions where the N1 form was administered (most parts of Canada, see section 3.1.1). In these regions, a systematic sample of approximately one in three households was drawn in the first phase. On July 14, 2011, those who had already responded formed the group known as the "initial respondents". Non-respondents as of that date formed the "initial non-respondents" group. A subsample of about one in three households of these initial non-respondents (for a combined sample fraction of about one in nine) was selected for non-response follow-up (NRFU). About 40% of these individuals responded during NRFU. As a result, weights associated with NHS respondents (or more simply "NHS weights") in N1 regions vary widely. In contrast, in regions where the N2 questionnaire was used (the N2 was administered to all households in these regions with no NRFU subsample), only non-response (and not sampling) affected the NHS weights. Thus, even with a response rate of 50% in a particular N2 region, the NHS weight would be only 2. A more detailed description of the NHS sampling design is found in Chapter 3 of the National Household Survey User Guide.

One of the assumptions used in the APS sample allocation method was that the NHS weights would vary as little as possible within an APS stratum. To this end, only part of the APS stratification came from the NHS stratification. Since the entire NHS stratification could not be used, it was necessary to select the stratification variables that caused the greatest variation possible in weights between the strata.

As previously mentioned, the type of region (N1 or N2) associated with the type of NHS form and the collection method created a significant difference in NHS weights. An optimum allocation between N1 and N2 regions would have been to select proportionately more individuals in the APS from N1 regions than from N2 regions in order to balance the weights combining the NHS and APS sampling.

Another factor causing differences in the NHS weights in the N1 regions was whether or not the respondent was part of the non-response follow-up (NRFU) sample. The weights of NRFU respondents were on average 7.5 times higher than those of initial respondents (who were not part of NRFU). One possible way to have offset this effect would have been to select proportionately 7.5 times more individuals in the NRFU respondent strata than in the initial respondent strata. However, this type of approach could have had numerous disadvantages (such as not being able to select enough NRFU respondents and the fact that the probabilities of response to the APS are definitely less among NRFU respondents than among initial respondents). Certain compromises had to be made during sample allocation.

Another factor, unrelated to the variability of NHS weights, but one that was very important to consider as a stratification factor was the type of Aboriginal identification (or more simply "Aboriginal type") of the respondent selected from the NHS. This factor refers to the fact that the NHS respondent indicated either Aboriginal identity (ID) or Aboriginal ancestry-only (AO) in the NHS. These two groups have very different characteristics. Moreover, allocation must reflect the probability of each unit being part of the target population (having Aboriginal identity on the APS). Nationally, based on the 2006 APS data, individuals having identity on the NHS have a probability of about 88% of having identity on the APS, while individuals with ancestry-only on the NHS have a probability of about 32% of having identity on the APS.

Combining the factors type of region (N1 or N2), type of NHS respondent (initial respondent or NRFU respondent) and Aboriginal type (AO or ID) gave a maximum of six strata per domain of estimation:

- N1, initial respondent, AO
- N1, initial respondent, ID
- N1, NRFU respondent, AO
- N1, NRFU respondent, ID
- N2, AO
- N2, ID

For certain domains such as Status First Nations, there were only three possible strata because included individuals had Aboriginal identity by definition. Some strata may also have been empty, especially in the supplementary domains. For example, for a specific Inuit region, it was possible that no one classified as Aboriginal ancestry-only had Aboriginal ancestry other than Inuit.

3.2.3 Sampling design and sample allocation

The 2011 APS selected its sample from 2011 NHS respondents. Section 3.2.2 described how the NHS used a two-phase sampling design in the N1 regions and a single-phase sampling design in the N2 regions. Thus, the APS sampling design can be considered a three-phase design in the N1 regions and a two-phase design in the N2 regions.

Once the frame had been constructed, it was stratified according to domain of estimation and then substratified by type of region (N1 vs. N2), type of NHS respondent (initial respondent vs. NRFU respondent) and Aboriginal type (ID vs. AO). A systematic random sample was then selected within each stratum, the frame having been pre-sorted by household and person number. The purpose of this was to ensure proper distribution of the sample geographically within the strata as well as across as many households as possible. This procedure was, however, limited by the fact that members of the same household can be in different strata.

A method for optimal allocation between the substrata of a particular domain was used by taking into account the different types of sample size loss as described in section 3.2.4 as well as the probability of each unit belonging to the target population in a given stratum. This allocation depended in part on the NHS weights. It should be noted that at the time of the allocation, these weights had not yet been calculated. Preliminary weights were therefore derived solely for the allocation. The definitive weights derived by the methodology team working on the NHS estimates were used during weighting (see section 6). For a reference on this topic, see Verret (2013).[2]

Allocation was done in the survey's target domains first. For the non-targeted or supplementary domains (for example, non-Inuit in Inuit regions or persons 45 and over among school completers and leavers), the sample size in a given stratum was calculated using a sampling fraction equal to or less than the sampling fraction of the corresponding stratum in the corresponding target domain. These supplementary domains had to be retained in order to cover the entire target population but did not require a specific sample size since no estimates were to be derived for them. For non-targeted Aboriginal groups, such as non-Inuit in Inuit regions, the same sampling fraction as for Inuit was used. For persons 45 and over, half of the sampling fraction of persons under 45 was used.

Although the plan was to select the 2012 APS sample in a single wave, after a few weeks of collection, it was decided to select a top-up sample (called the wave 2 sample). There were three reasons for this decision:

- Since the RDB (APS sampling frame) had not been fully edited at the time of APS sample selection, a number of individuals appeared as Status Indians or as members of First Nations / Indian bands only (see section 3.2.1). The initial APS sample contained 1,254 of these cases. These units will be referred as "false Aboriginal individuals" later in this guide;

2. Verret, F. (2013). The estimation methodology of the 2011 National Household Survey. Paper presented at the 2013 Joint Statistical Meeting.

- The response rate among school leavers was lower than for the other groups and also lower than expected;
- The proportion of the sample in Quebec belonging to the target population (rate of Aboriginal identity on the survey) was lower than expected for both those who had the identity on the NHS and those who had Aboriginal ancestry-only on the NHS.

To calculate the size of the top-up sample, the "false Aboriginal individuals" were first removed from the survey frame. The response rates among school leavers were adjusted downward and the identity rates on the survey were reviewed for Quebec based on the results observed at the start of collection. The same allocation method was then used to determine new sample sizes. The sample size obtained using this method represents the size that would have been used if the "false Aboriginal individuals" had never existed on the frame and if it had been possible to more precisely estimate the response rates of school leavers and identity rates in Quebec. The difference between this sample size and the size of the initial sample constitutes the size of the top-up sample. No additional units were selected in the supplementary domains. After removing the "false Aboriginal individuals", the initial sample contained 49,287 units and the top-up sample 1,737 units, for a total of 51,024 units.

3.2.4 Sample size adjustment

a) Probability of belonging to the target population

The 2006 APS found that 88% of individuals with Aboriginal identity in the 2006 Census reported Aboriginal identity in the 2006 APS and 32% of those with ancestry-only in the 2006 Census reported identity in the 2006 APS. As a result of changes in the survey methodology, questionnaire and in social environment, these probabilities of belonging to the target population were expected to be lower in 2012 than in 2006. Accordingly, the probabilities of belonging to the specified 2012 target population were adjusted downward by multiplying 2006 figures by 75% in the NHS Aboriginal ancestry-only strata and by 95% in the NHS identity strata.

b) Response rates

Response rates from the 2006 APS were used to allocate the sample for the 2012 APS. As a preventive measure, these rates were reduced by multiplying them by an adjustment factor of 90%. Note that increasing the sample size based on an expected response rate is a protective measure to reduce sampling error but does nothing to reduce non-response bias, a particular type of non-sampling error (see section 7.3). Clearly, it is preferable to obtain a higher response rate on a smaller sample size than a lower response rate on a larger sample size.

c) Transition between educational groups

Estimates were targeted for four education groups (or sometimes two) per region and Aboriginal group. However, the classification of individuals among these groups was done using education and age variables, which can sometimes be inaccurately reported or missing on the NHS. In many cases, and for children in particular, educational attainment could only be approximated using NHS data (in fact, only date of birth was used to determine the education group of children under 15 years of age). As well, changes over time meant that NHS data could not allow for a prediction of the exact educational group of APS respondents as of February 1, 2012, the survey's reference date.

Loss factors related to transitions between groups were estimated for the 2012 APS, always using 2006 data. For 2006 APS respondents, education groups were derived based on the 2006 Census RDB data and the categorizations were compared to the education groups obtained during the 2006 APS. The 2006 APS results had shown slightly more individuals than expected attending elementary school (2.5% more), fewer individuals than expected attending high school (8.5% less), slightly fewer school completers than expected (5.8% less), and more school leavers than expected (15% more). Size adjustments factors were calculated by region, Aboriginal group and education group. These factors were multiplied by 1.05 as a precaution

measure. If a factor was still smaller than 1, the factor was simply raised to 1 so as not to reduce sample sizes. For example, even though more school leavers were expected on the APS than on the NHS, the size of the leavers group was not reduced as a precaution. These factors were included in the allocation process in order to increase (or not) the sample size accordingly.

d) Households with more than three individuals selected

Because the sample unit was the individual and not the household, the sample may have included several individuals from a single household. To limit response burden, one of the constraints imposed during collection was to select no more than three individuals per household. As a way to estimate this loss, a preliminary sample was selected and the number of persons who would have been removed from the sample because of this constraint was calculated. The relative loss obtained was multiplied by 1.1 for each combination of domains and strata and was incorporated in the sample size adjustment factor.

e) Overlap with other surveys

The APS collection period overlapped with that of several other Statistics Canada surveys. It is generally understood that an individual contacted for one survey is less likely to agree to participate in that survey if he or she has just been interviewed for another survey. To compensate for this potential loss, the total sample size was increased uniformly by about 5%. Several of the overlapping surveys also cover the territories where the Aboriginal population represents a high portion of the total population. Therefore, there was a higher possibility of overlap between the different surveys in the territories. Special attention was given to this matter in order to reduce overlap as much as possible and, implicitly, to reduce the potential burden on respondents. Overlap in the territories was examined for the Canadian Community Health Survey (CCHS), the Labour Force Survey (LFS) and the Programme for the International Assessment of Adult Competencies (PIAAC) and steps were taken to minimize response burden where possible.

For example, overlap with the CCHS was of particular concern. Some of the APS questions on health are drawn directly from the CCHS survey, including several highly sensitive questions on mental health. No measures were taken for the CCHS interviews held before and after the APS collection, that is, before January 2012 and after June 2012. However, the identifiable overlap in households to be interviewed between January 2012 and June 2012 for the CCHS was eliminated. It should be mentioned that the area frame of the LFS (from which the CCHS sample is selected) had been matched with the Census resulting in approximately 70% of the households of the area frame being matched to the Census. Since the CCHS selects a systematic random sample of households within each community covered, households not drawn also form a systematic random sample of households. Households selected for the CCHS between January and June and which had responded to the NHS were removed from the APS survey frame. Within each community covered by the CCHS, the weights of the remaining NHS households were adjusted to represent all of the community's households. Since these households had been removed from the APS survey frame, they therefore had no chance of being selected.

3.3 Sample size and response rate

The initial and top-up sample of the 2012 APS contained a total of 51,024 units, 376 of which were removed before being sent to collection (see section 6.2). Table 4 shows the allocation of the 50,648 units sampled by geographical domain and the corresponding response rates.

A response rate is defined as the number of eligible respondents divided by the number of eligible units in the sample. Persons living outside of Canada or in an institution at the time of the survey or under 6 years of age as of February 1, 2012 are examples of ineligible units for the APS.

Two definitions of eligible units were used for the APS. In the first definition, individuals without Aboriginal identity on the APS were deemed ineligible (in addition to the other types of ineligible units). In the second definition, these individuals were deemed eligible and were included as respondents. Indeed, these individuals agreed to participate in the survey and completed all of the questions administered to them, that is, the screening questions determining whether or not they have Aboriginal identity.

These two definitions of eligible units and respondents therefore offer two ways to measure the response rate. The first response rate of 70.4% is generally the response rate used during collection. This is a response rate relative to the sampled units falling within the APS identity population.The second response rate of 76.2% is the one used by Methodology and is more a response rate relative to all sampled units (with some exclusions, however). Of course, in both cases, non-responding units cannot always be classified as being eligible or not eligible. Given that the probability of having identity on the APS differs substantially between individuals with Aboriginal identity on the NHS and those with ancestry-only on the NHS, the response rate used during collection is highly influenced by the allocation of the sample between these two groups, which is not the case for the response rate used by Methodology. In Inuit regions, because most individuals have identity, the two response rates are quite similar, which is not the case in the other regions. The maximum difference between the two rates occurs in Quebec outside Nunavik. In this region, the Aboriginal ancestry-only population forms a large part of the total Aboriginal population and, according to the 2006 APS, the probabilities of having Aboriginal identity on the survey are the lowest in the country regardless of whether the individual has Aboriginal ancestry-only (20%) or Aboriginal identity on the NHS (66%). Therefore, although the methodology response rate is the highest in Quebec, the corresponding data collection response rate is relatively low.

Table 4 Sample size and response rate by geographical domain using two definitions

	Total	Eligible 1	Eligible 2	Respondents 1	Respondents 2	Response rate 1	Response rate 2
Geographical domain	number of units					percentage	
Nunatsiavut	777	752	759	604	611	80.3	80.5
Nunavik	2,084	2,000	2,021	1,543	1,564	77.2	77.4
Inuvialuit	1,153	1,117	1,139	791	813	70.8	71.4
Nunavut	2,181	2,128	2,156	1,581	1,609	74.3	74.6
Nunangat (total)	**6,195**	**5,997**	**6,075**	**4,519**	**4,597**	**75.4**	**75.7**
Atlantic excluding Nunatsiavut	3,352	2,640	3,336	2,025	2,721	76.7	81.6
Quebec excluding Nunavik	6,776	3,155	6,724	2,317	5,886	73.4	87.5
Ontario	8,366	6,299	8,306	4,286	6,293	68.0	75.8
Manitoba	5,190	4,589	5,136	3,006	3,553	65.5	69.2
Saskatchewan	4,946	4,412	4,880	3,124	3,572	70.8	73.5
Alberta	6,603	5,477	6,547	3,765	4,835	68.7	73.9
British Columbia	6,963	5,707	6,901	3,918	5,112	68.7	74.1
Yukon	786	712	764	489	541	68.7	70.8
Northwest Territories excluding Inuvialuit	1,471	1,365	1,440	960	1,035	70.3	71.9
Rest of Canada (total)	**44,453**	**34,356**	**44,014**	**23,890**	**33,548**	**69.5**	**76.2**
Canada (total)	**50,648**	**40,353**	**50,089**	**28,409**	**38,145**	**70.4**	**76.2**

Eligible 1 Sample size excluding out of scope on the APS (non-Aboriginal individuals are out of cope under this definition)
Eligible 2 Sample size excluding out of scope on the APS (non-Aboriginal individuals are in scope under this definition)
Respondents 1 Number of respondents excluding non-Aboriginal individuals on the APS
Respondents 2 Number of respondents including non-Aboriginal individuals on the APS
Response rate 1 Response rate excluding non-Aboriginal individuals on the APS
Response rate 2 Response rate including non-Aboriginal individuals on the APS

4. Data collection

4.1 Time frame

The 2012 Aboriginal Peoples Survey was conducted from February 6, 2012 to July 30, 2012. As a postcensal/post-NHS survey, it followed the 2011 National Household Survey which was conducted between May and August 2011.

4.2 Mode of collection

The questions in the 2012 APS were administered in a computer assisted interviewing (CAI) environment. Two computer-assisted interview methods were used for this survey: Computer Assisted Telephone Interviews (CATI) and Computer Assisted Personal Interviews (CAPI). In most regions, CATI was used for individuals for whom there was a telephone number on the sample file. CAPI was used for individuals who did not have a telephone number or who could not be contacted by telephone even when a number was available. In the territories, the northern parts of many provinces and some Inuit communities, there were often very few telephone numbers available. In these cases, personal interviews were conducted.

Respondents were interviewed in the official language of their choice. For Inuit regions, the questionnaire was translated as a paper copy into Inuktitut (Baffin dialect) and an Inuktitut audio recording of the questionnaire was made to assist interviewers with potential language barriers in the field.

The time required to complete the survey varied from person to person. In some cases, the 2012 APS interview took up to an hour or more to finish, but on average the survey took about 40 minutes to complete.

4.3 Supervision and quality control

All Statistics Canada interviewers were under the supervision of senior interviewers who were responsible for ensuring that interviewers were familiar with the concepts and procedures of the surveys to which they were assigned. Senior interviewers were also responsible for periodically monitoring the interviewers.

Interviewers were trained on the survey content and the computer-assisted interviewing application. In addition to classroom training, the interviewers completed a series of mock interviews to become familiar with the survey and its concepts and definitions.

4.4 Proxy interviews

Questions for selected children and youth (aged 6 to 14) were completed by proxy interview, that is, by the Person Most Knowledgeable (PMK) about the selected person. The PMK was generally the parent or guardian of the child. For individuals between the ages of 15 and 17, interviews were conducted directly with the youth only with the prior approval of the individual's parent or guardian. Where approval was not provided, the data were collected from the parent or guardian.

The adult questionnaires were designed to be answered by the selected person him/herself. Proxy interviews were acceptable in some circumstances, such as when the selected respondent was not able to answer due to mental or physical health, due to a language barrier, or because the selected respondent was absent from home for the duration of the survey.

4.5 Communications strategy

In the months leading up to data collection, efforts were made to raise awareness of the 2012 Aboriginal Peoples Survey and to encourage participation. The communications strategy included the development of an APS brochure and posters. Statistics Canada's team of 11 Aboriginal Liaison Program Advisors served as contacts for the APS in regions across the country. They put up posters, distributed brochures and circulated information about the survey by way of e-mail, newsletters and websites. They met with local and provincial organizations and Friendship Centres across Canada. All 119 Friendship Centres displayed APS posters and brochures and kept their websites updated on APS survey information. The Métis National Council (MNC) and the Congress of Aboriginal Peoples (CAP) also helped to promote the survey.

Prior to collection, introductory letters and brochures were mailed to selected respondents outlining the purpose of the survey and emphasizing the importance of their participation. Additional Inuktitut/English introductory letters were prepared and printed for hand delivery to houses in the North which could not otherwise be reached. Public Service Announcements were prepared for broadcast over local radio stations. They briefly announced the arrival of the survey in the community and explained that a Statistics Canada interviewer could be coming to their door. These announcements were translated into the four main Inuit languages in each of the four Inuit regions.

Once collection was complete, follow up thank you letters were sent to respondents who participated in the survey along with an APS calendar.

4.6 Special issues in the field

Interviewers were instructed to make all reasonable attempts to obtain a completed interview with the selected member of the household. Those who at first refused to participate were re-contacted up to two more times to explain the importance of the survey and to encourage their participation. For cases in which the timing of the interviewer's call was inconvenient, an appointment was arranged to call back at a more convenient time. For cases in which there was no one home, numerous call backs were made.

Special issues arose in relation to data collection for the APS which were addressed with extra coordination in the field and corrective adjustments to survey methods. For instance, the collection of the 2012 APS occurred during the same time period as several other surveys conducted by Statistics Canada, creating a potentially heavy response burden for individuals who might be selected for participation in more than one survey. Careful planning and adjustments to survey design were implemented to address and minimize this type of respondent-burden (please refer to section 3.2.4(e) for more information).

4.7 Final response rates

Approximately 38,150 respondents completed the 2012 Aboriginal Peoples Survey including those who reported being non-Aboriginal on the APS for a final response rate of 76%. Excluding approximately 9,740 non-Aboriginal respondents, the total number of Aboriginal respondents included in the 2012 APS database was about 28,410.

A detailed description of the sample design used and response rates obtained is provided in Chapter 3 of this guide. Table 4 in section 3.3 provides the final response rates and sample sizes for each of the geographic domains covered by the survey (e.g. provinces, territories).

5. Data processing

5.1 Data capture

Responses to survey questions were captured directly by the interviewer at the time of the interview using a computerized questionnaire. A computerized questionnaire reduces processing time and costs associated with data entry, transcription errors and data transmission.

Some editing of data was done directly at the time of the interview. Specifically, where a particular response appeared to be inconsistent with previous answers or outside of expected values, the interviewer was prompted, through message screens on the computer, to confirm answers with the respondent, and, if needed, to modify the information.

5.2 Social survey processing steps

Data processing involves a series of steps to convert the electronic questionnaire responses from their initial raw format to a high-quality, user-friendly database involving a comprehensive set of variables for analysis. A series of data operations are executed to clean files of inadvertent errors, rigorously edit the data for consistency, code open-ended questions, create useful variables for data analysis, and finally to systematize and document the variables for ease of analytical usage.

The 2012 APS used a new set of social survey processing tools developed at Statistics Canada called the "Social Survey Processing Environment" (SSPE). The SSPE involves SAS software programs, custom applications and manual processes for performing the following systematic steps:

Processing steps:

- Receipt of raw data
- Clean up
- Recodes
- Flows
- Coding
- Edits and imputations
- Derived variables
- Creation of final processing file
- Creation of dissemination files

5.3 Receipt of raw data and record clean up

Following the receipt of raw data from the electronic questionnaire applications, a number of preliminary cleaning procedures were implemented for the 2012 APS at the level of individual records. These included the removal of all personal identifier information from the files, such as names and addresses, as part of a rigorous set of ongoing mechanisms for protecting the confidentiality of respondents. Duplicate records were resolved at this stage. Also part of clean up procedures was the review of all respondent records to ensure each respondent was "in-scope" and had a sufficiently completed questionnaire. (Note that the criteria to determine whether or not a respondent was in scope was applied before any edit or imputation was done). Specific criteria for determining who would be a final APS respondent and who would not be a final APS respondent are provided below.

5.3.1 Definition of a respondent

1. To be "in scope", respondents must have been at least 6 years of age as of February 1, 2012 and met a minimum of one Aboriginal identity criterion (see section 2.2 for complete criteria).
2. To have a "complete" questionnaire, respondents aged 6 to 14 must have provided valid responses (i.e. not "Don't know" or "Refused") to specified key questions in the areas of education or health.
3. To have a "complete" questionnaire, respondents aged 15 and older must have provided valid responses (i.e. not "Don't know" or "Refused") to specified key questions in either the area of education, or the areas of labour **and** health.

Those that did not meet the above criteria were removed from the database. As per the rules above, all "partial" respondents, who were in-scope according to part 1 of the definition but who did not fulfill the content-completion requirements in part 2 or part 3 of the definition, were among those removed from the final database. Please refer to section 6.4 of this document for more information on partial respondents.

5.4 Variable recodes and multiple response questions

This stage of processing involved changes at the level of individual variables. Variables could be dropped, recoded, re-sized or left as is. Formatting changes were intended to facilitate processing as well as analysis of the data by end-users. One such change was the conversion of multiple-response questions ("Mark-all-that-apply" questions) to corresponding sets of single-response variables which are easier to use. For each response category associated with the original question, a variable was created with yes/no response values. An example is provided below.

Original multiple-response question:

ED4_Q11A - What were the reasons you did not finish your postsecondary education?

INTERVIEWER: Mark all that apply.

01	Pregnant/Caring for own child(ren)
02	Other family responsibilities
03	Own illness / Disability
04	Financial reasons (not enough money)
05	Lost interest / Lack of motivation
06	Got a job / Wanted to work
07	Too old or too late now
08	Courses too hard / Bad results
09	Too difficult to be away from home
10	Prejudice and racism
11	Moved
12	Other - Specify
DK, RF	

Final variables in single-response yes/no format:

ED4_Q11AA - What were the reasons you did not finish your postsecondary education?

- Pregnant/Caring for own child(ren)

1	Yes
2	No
DK, RF	

ED4_Q11AB - What were the reasons you did not finish your postsecondary education?

- Other family responsibilities

1	Yes
2	No
DK, RF	

ED4_Q11AC - What were the reasons you did not finish your postsecondary education?

- Own illness / Disability

1	Yes
2	No
DK, RF	

...additional Yes-No questions for each response category, as indicated, from "Financial reasons (not enough money)" to "Moved"... and including the last category:

ED4_Q11AL - What were the reasons you did not finish your postsecondary education?

- Other - Specify

1	Yes
2	No
DK, RF	

5.5 Flows: response paths, valid skips and question non-response

Another set of data processing procedures for the 2012 APS was the verification of questionnaire flows or skip patterns. All response paths and skip patterns that were built into the questionnaire were verified to ensure that the universe or target population for each question was accurately captured during processing. Special attention was paid to distinctions between valid skips and non-response, an important distinction for statistical analysis. These concepts are explained below in order to assist users to better understand question universes as well as statistical outputs for APS survey variables.

Response – an answer directly relevant to the content of the question that can be categorized into pre-existing answer categories, including "Other-specify".

Valid skip – indicates that the question was skipped because it did not apply to the respondent's situation, as determined by valid answers to a previous questions. In such cases, the respondent is not considered to be part of the target population or universe for that question. As noted below, where a question was skipped due to an undetermined path (that is, a "Don't know" or "Refusal" to a previous question caused the skip), the respondent is coded to "Not stated" for that question.

Don't know – the respondent was unable to provide a response for one or more reasons (due to lack of recall, or because they were responding for someone else, for example).

Refusal – the respondent refused to respond, perhaps due to the sensitivity of the question.

Not stated – this indicates that the question response is missing and there is an undetermined path for the respondent, such as when a respondent did not answer the previous filter question or where an inconsistency was found in a series of responses.

Special codes have been designated to each of these types of responses to facilitate user recognition and data analysis. For instance, "valid skip" codes are set to "6" as the last digit, with any preceding digits set to "9" (for example, code would be "996" for a 3-digit variable). All "Don't know" responses end in "7", with any preceding digits set to "9" (for example, "997"). Refusals end in "8", with any preceding digits set to "9" (for example, "998"); and "Not stated" values end in 9, with any preceding digits set to "9" also (for example, "999").

5.6 Coding

5.6.1 "Other-specify" items

Data processing also includes the coding of "Other-specify" items, also referred to as "write-in responses". For most questions on the APS questionnaire, pre-coded answer categories were supplied and the interviewers were trained to assign a respondent's answers to the appropriate category. However, in the event that a respondent's answer could not be easily assigned to an existing category, many questions also allowed the interviewer to enter a long-answer text response in the "Other-specify" category.

All questions with "Other-specify" categories were closely examined during processing. Based on a qualitative review of the types of text responses given, coding guidelines were developed for each question. Based on these coding guidelines, many of the long answers provided were re-coded back into one of the pre-

existing listed categories. Responses that were unique and different from existing categories were kept as "Other". For some questions, one or more new categories were created when there were sufficient numbers of responses to warrant them. In the case of questions where "Other-specify" responses constituted less than 5% of overall responses to the question, coding was not performed and responses were left in "Other".

Approximately 58,000 responses across 78 questionnaire items were recorded under "Other – specify" and reviewed for coding. Appendix B summarizes the extra categories added for the 2012 APS. These will be taken into account when refining the answer categories for future cycles of the survey.

5.6.2 Open-ended questions and standard classifications

A few questions on the 2012 APS questionnaire were recorded by interviewers in a completely open-ended format. These included questions related to the respondent's occupation and industry of work as well as their major field of post-secondary study, where applicable. These responses were coded using a combination of automated and interactive coding procedures. Standardized classification systems were used to code these responses. Appendix C provides details of these classifications.

A standardized classification was also used to code Aboriginal languages that respondents spoke or understood as well as the first language learned in childhood. For languages, interviewers had been provided a comprehensive drop-down menu of languages to choose from, but write-in responses were also captured as needed. Overall, 51 Aboriginal language categories were used to code APS language data. For details on the classification system used for Aboriginal languages, see Appendix C.

Coding for all classifications involved experienced coding and quality control as well as additional processing verification procedures.

5.7 Edit and imputation

After the coding stage of processing, a series of customized edits were performed on the data. These consisted of validity checks within and across variables to identify gaps, inconsistencies, extreme outliers and other problems in the data. To resolve the problematic data identified by the edits, corrections were performed based on logical edit rules. In some cases, corresponding data were taken from the respondent's answers to the National Household Survey. This is referred to as imputation.

An example of a validity check within a single question is the housing variable related to the number of rooms in the dwelling, which allowed for an interviewer to record up to 95 rooms. To remove outlier responses that were suspected of being invalid, an edit was built to ensure that the reported number of rooms in the dwelling did not exceed an upper limit of 20. As another example, many consistency edits across questions were performed in relation to education variables to avoid any contradictions in education profiles. For example, a person who had not reported ever having attended a specific post-secondary educational institution such as a university, a trade school, a college, CEGEP or other non-university institution, and then subsequently reported currently working toward a certificate, diploma or degree from one of these institutions, was assumed to have attended that type of institution. The response to the earlier question was changed from a "no" to a "yes" for the specific type of institution where the edit was required.

For the 2012 APS, a series of important imputations was conducted in relation to Aboriginal identity classifications. For example, those with missing data for questions ID_Q02 on Aboriginal identity group, ID_Q03 on Registered Indian Status, or ID_Q05 on membership in a First Nation or Indian band were imputed values based on their responses to the National Household Survey. For those who self-reported as an Aboriginal person on APS question ID_Q01 but who did not report any specific Aboriginal group in ID_Q02, an imputation was also conducted based on the respondent's answer to the NHS. In addition, an imputation was performed for a person who had not identified being in any Aboriginal group but had identified as being either (1) a Status Indian, (2) registered as a Status Indian under Bill C-31 or Bill C-3, or (3) a member of a First Nation or Indian band – these respondents were imputed to First Nations people (North American Indian).

Finally, although all of these edits across topics were performed systematically using computer programmed edits, there were some cases for which very complex combinations of information were reviewed and corrected manually.

5.8 Derived variables and NHS linkage

In order to facilitate more in-depth analysis of the rich APS dataset, over 500 derived variables were created by combining items on the questionnaire. Derived variables (DVs) were created across all major content domains. In addition, more than 100 National Household Survey variables were linked to the final APS analytical file for 2012.

Many of the derived variables were straightforward and involved simply combining equivalent questions, such as those across educational streams, for instance. Other simple derived variables involved the collapsing of categories into broader categories. In other cases, two or more variables were combined to create a new or more complex variable which would be useful for data analysts. Some of the derived variables were based on linked variables from the NHS, including multiple NHS geographies and Inuit regions. Aboriginal ancestry was also taken from the NHS since it is not measured directly by the 2012 APS.

In constructing derived variables, a valid response category was generally not assigned to a respondent for a given derived variable if any part of the equation was not answered (that is, if any question used in the derived variable had been coded to "Don't know", "Refused" or "Not stated"). In such cases, the code assigned to the derived variable was labelled "Not stated".

Most derived variable names have a "D" in the first character position of the name. Geography DVs are the exception, since they reflect the corresponding NHS variable name. For all linked NHS variables, the NHS variable name was preserved as much as possible on the APS database. Some exceptions applied since APS variable names are restricted to eight characters whereas NHS variable names sometimes exceeded eight characters.

The 2012 APS codebook (data dictionary) identifies in detail which variables were derived and provides information on how the derivations were done. Highlights of DVs are listed by theme in Appendix A along with other survey indicators. A complete list of linked NHS variables and their accompanying notes are provided in the 2012 APS codebook (data dictionary) which accompanies the APS analytical file.

5.9 Creation of final data files and codebook (data dictionary)

Four final data files were created in data processing:

- Final processing file
- Analytical file for use in Research Data Centres
- Public use microdata file (PUMF)
- Inuit share files, as per data sharing agreement with the four Inuit regions

The final processing file is an in-house file that includes a number of temporary variables used exclusively for processing purposes. The Analytical File, the Public use microdata file (PUMF) and the Inuit share files are dissemination files which are processed further for release purposes. Dissemination files are scheduled for distribution at various points in time following the APS release day of November 25, 2013 (please refer to Chapter 9 for more detailed descriptions and dissemination details).

The analytical file is distributed in Research Data Centres across Canada but can only be accessed by researchers who fulfill certain requirements. The analytical file is also used at Statistics Canada to produce data tables in response to client requests. The PUMF is constructed for wider public distribution. The Inuit share files are produced in accordance with data sharing agreements with the Inuit regions: Nunatsiavut, Nunavik, Nunavut and the Inuvialuit region. On all of these dissemination files, many steps have been taken to ensure respondent confidentiality.

In order to transform the final, cleaned processing file to a final Analytical file for researchers, a number of steps were taken. First, a series of steps were taken for the enhanced protection of respondent confidentiality. Next, person-weights were added to the file. Weighting is described in more detail in chapter 6. Finally, all temporary variables or variables used exclusively for processing purposes were removed from the final processing file.

Accompanying the 2012 APS analytical file is the record layout, SAS (Statistical Analysis System) and SPSS (Statistical Package for the Social Sciences) syntax to load the file, and metadata in the form of a codebook that describes each variable and provides weighted and unweighted frequency counts. The codebook is also referred to as the data dictionary.

The public use microdata file (PUMF) undergoes more extensive data processing for the protection of respondent confidentiality. In order to ensure the non-disclosure of confidential information, the level of detail of the PUMF is not as fine as that of the analytical files kept by Statistics Canada. Actions are taken to protect against the recognition of respondents with potentially identifiable combinations of characteristics. These protective actions include the restriction of geographies included in the file, adjustments to survey weights, review of overlaps with other PUMFs being published, exclusion of variables, grouping of categories for some variables, capping of some extreme numerical values, as well as identification of unique records at risk and rare occurrences.

6. Weighting

To calculate initial weights, "false Aboriginal individuals" were treated as though they never existed, that is, they were excluded from the survey frame and the sample (please refer to section 3.2.3 "Sampling design and allocation of the sample" for more information).

In a sample survey, each selected person represents not only himself or herself, but also other persons who were not sampled. Consequently, a weight is associated with each selected person to indicate the number of persons that he or she represents. This weight must be used for all estimations. For example, in a simple random sample of 2% of the population, each person represents 50 persons in the population. The initial weight is then adjusted for such things as non-response and discrepancies between the characteristics of the sample and known totals for the target population (post-stratification). In fact, seven steps were used in the weighting process.

6.1 Initial weights

The initial weight of a unit in a given APS stratum corresponds to the product of two components: the inverse of the stratum sampling fraction and the NHS weight corrected for non-response to the NHS for the unit in question. The stratum sampling fraction is calculated as the number of people selected for the APS in each stratum divided by the total number of available NHS respondents for that stratum. The NHS weight used is the NHS sampling weight corrected for non-response, then capped to the 99th percentile, as calculated by the methodology team working on NHS estimation.

6.2 Adjustment for units not sent to collection

A relatively small number of sampled units were not sent to the field for different reasons. These included:

- cases where three members of the same household had already been selected;
- units without a name or date of birth;
- "wave 2 ineligible units", that is, individuals selected in wave 2 in households where at least one individual had indicated his refusal to participate in the survey in wave 1.

In the first two instances, a ratio adjustment was made by NHS region and Aboriginal group. In the third instance, a ratio adjustment was made by NHS region, Aboriginal group and education group. Within a region and Aboriginal group (or a region, Aboriginal group and education group in the third case), the weights of units removed were set to zero and the weights of the remaining units were increased proportionally (ratio adjustment).

6.3 Adjustment for non-response

Two adjustments were made for two types of non-response: the selected person for whom no contact was made or the parent or guardian of the child ("non-contact": 2,981 adults and 770 children) and persons contacted who did not (or could not) provide the information for themselves or their child ("non-response with contact": 6,263 adults and 1,763 children). The second type of non-response is mainly associated with refusals or "disguised refusals". An example of a "disguised refusal" might be a person contacted several times who continually postpones the interview. Two adjustments were made since the characteristics of the people that could not be contacted are often different from those of the people who refused when contacted.

The distinction between children and adults is made here based on age according to the NHS (and not age as of February 1, 2012 as measured on the APS), that is, under 15 years of age for children and 15 and over for adults. This is an important distinction because fewer characteristics explaining non-response are available for children than for adults on the NHS. Among children, it is not the characteristics of the child that influence response or non-response but rather the characteristics of the person responding for the child (parent or guardian). Consequently, it was necessary to determine for each child under 15 on the NHS who the most likely person was to respond for the child based on the child's situation in the census family, and regardless of whether a response was or was not obtained to the APS for this child. In situations where the child's parents or guardians lived as opposite sex couples, preference was given to the female person (mother, grandmother, aunt, for example).

It should be mentioned that the definition of "non-contact" changed from the definition used at the time of the 2006 APS. Because the 2012 interviews were computer-assisted interviews (CAI) rather than the paper format questionnaire used in 2006, a series of collection variables, referred to as "paradata" were available for all units of the sample. In particular, information was collected for each contact attempt. A unit was deemed "non-contact" if none of the attempts resulted in contacting the person selected or the parent or guardian of the child selected. In 2006, non-contact was established based on the last contact attempt only. Consequently, in 2012 compared to 2006, there were proportionally fewer "non-contacts" and more "non-responses with contact".

Weights were first adjusted for non-contact cases and then for non-response with contact, for adults and children separately. In what follows, the term "non-response" will be used for both types of non-response. The term "respondent" refers to the person completing the information for the selected person (usually themselves for the adults or a parent or guardian for the children).

Each non-response adjustment was done in three steps. First, a logistic regression model was used to predict the response probability (probability of obtaining a response) for each selected unit (for both responding and non-responding units) from a series of explanatory variables. These explanatory variables are divided into two groups. The first group consists of the "person" or "household" characteristics from the NHS for the person selected or of the parent or guardian of the child selected (for example, Aboriginal group of the person selected or of the parent or guardian of the child selected, number of people in the household of the person selected, etc.). The second group of explanatory variables consists of collection variables called "paradata". The number of attempts to contact a subject and whether tracing was required are examples of paradata variables used by logistic regression models. The paradata were found to be particularly good predictors of the response or non-response as many of these variables measure the effort to contact a person or to obtain a response from a contacted person. For instance, individuals requiring a large number of attempts to be contacted were found to be very similar to individuals for whom no contact was made (all attempts failed).

In the second step, individuals (respondents and non-respondents) with similar response probabilities were grouped in adjustment classes using cluster analysis. A simulation was carried out to determine approximately the optimal number of classes and the minimum number of respondents per class. The response rate was derived for each class based on the number of respondents and non-respondents in the class. The derived response rate was weighted using the weights from the previous adjustment step.

In the third step, the inverse of the weighted response rate in a class was used as the adjustment factor for that class and the weights of the responding units within the class were adjusted accordingly. The weights of the non-responding units were set to zero.

It is important to note that at this stage, all units considered to be out of scope were classified as respondents. Indeed, all the required information was collected from these individuals to determine that they were out of scope. The weights of these out-of-scope units were set to 0 in the last step of the weight adjustment and these units were eliminated from the analytical file. Retaining them until the last step makes it possible to produce internally weighted estimates of different groups of units outside the target population. This will be very useful, for example, in estimating certain parameters at the time of the next survey.

6.4 Adjustment for partial respondents

Partial respondents are individuals with Aboriginal identity in the APS but who did not complete enough information to meet the definition of respondent as defined in section 5. There were 157 partial respondents, which means their impact on the estimates should be minimal.

The adjustment was made by region, Aboriginal group and education group as measured on the NHS. A number of groupings were made by cross-tabulating these variables in order to obtain enough observations to calculate the adjustment factor. Knowing that these partial respondents had reported Aboriginal identity, only the weights for respondents of Aboriginal identity were increased to reflect partial respondents (the out-of-scope weights, including non-Aboriginal individuals on the APS, were not adjusted). The weights of partial respondents were then set at zero.

6.5 Post-stratification

Post-stratification ensures that the sum of the adjusted weights for the responding units corresponds to the NHS estimates according to different groups called post-strata.

In the case of the APS, two separate post-stratifications were carried out. The first post-stratification adjusts the weights of the Aboriginal identity or ancestry population from the NHS by post-stratum using the identity and ancestry variables from the RDB survey frame (see section 3.1.3) at the time of sample selection (and not the APS-measured variables, which are the subject of the second post-stratification). The post-strata are defined from certain combinations of region, Aboriginal type (identity or ancestry-only), Aboriginal group (Status First Nations, Non-Status First Nations, Métis, Inuit, other) and age group (6-14, 15-44, and 45 and over). The distinction between Status and Non-Status First Nations was used only for the provinces between Ontario and British Columbia. It is important to point out that the NHS estimates on which the weights were adjusted correspond exactly to the APS coverage, specifically, the identity or ancestry-only population aged 6 and over as of February 1, 2012, excluding people living on reserves and certain First Nations communities in the territories.

The weights were adjusted according to the ratio of the NHS weighted estimate to the sample weighted estimate for each post-stratum. As a result, the sample did not under- or over-represent certain combinations of Aboriginal groups, regions and age groups of the NHS.

Given that the responses to the questions defining the Aboriginal identity population (presented in section 3.1.1) may differ between the APS and NHS, a second post-stratification was carried out. Note that the APS questions defining the identity population are slightly different from those asked in the NHS (see Table 1 in section 2 and section 3.1.1). The second post-stratification ensured that the Aboriginal identity population estimated from the APS questions corresponded to the Aboriginal identity population defined according to the NHS within each post-stratum. Unlike the first post-stratification, the second one was not a "classical" post-stratification where weights were readjusted to address under- or over-representation of certain groups in the sample. Indeed, the answers to the questions on Aboriginal identity in the APS may have differed from those obtained by the NHS for a variety of reasons (section 8.1). This second post-stratification was more of a "practical" one that ensured that the Aboriginal identity population counts according to the APS were the same as those obtained on the NHS. After this step, only respondents with Aboriginal identity according to the APS had positive weights.

It is important to note that the 2012 APS processing and imputation system eliminated one category of Aboriginal identity, namely, the "Status Indian or member of a First Nation / Indian band only" group (see section 3.2.1). Respondents in this group were imputed as First Nations people in the 2012 survey. During the second post-stratification, individuals in this NHS group were also combined with First Nations people.

Because it was impossible to preserve the multiple identity counts between the APS and NHS (counts too small or discrepancies too large), individuals reporting an identity of First Nations and Métis, First Nations and Inuit, or First Nations, Métis and Inuit were combined with individuals reporting a First Nations identity during the second post-stratification. Individuals reporting a Métis and Inuit identity were combined with Métis. The second post-strata were formed from specific combinations of region, Aboriginal identity group (Status First Nations, Non-Status First Nations, Métis, Inuit) and age group (6-14, 15-44 and 45 and over).

6.6 Adjustment for extreme weights– Sigma gap method

Once the above weight adjustments were completed, some weights had very large values compared to others, which could have created problems during estimation if the observations with large weights also had very distinct characteristics from the observations with smaller weights. A method referred to as the "sigma gap" method was used to detect these extreme weights within each post-stratum, the post-strata being closely linked to the survey's domains of estimation (see section 3.2.1). Bernier and Nobrega (1998)[3] describe one application of the sigma gap method. The sigma gap method used here was intended to detect "outlier values" (excessively large weights) by calculating the difference between two successive weights after being sorted in descending order. This difference was compared to *n*standard deviation* of the weights within each post-stratum. If the difference exceeded *n*standard deviation* of the weights, the largest weight was identified as an outlier. Once a weight was identified as an outlier, then all others that were larger than it, in its post-stratum, were automatically identified as outliers. These weights are then reduced to the value of the first non-outlier weight. The mass of the reduced weights were then redistributed within the post-strata by a ratio adjustment. After examining a number of scenarios, a value of 2 was finally selected for *n*. This particular value for n made it possible to identify the weights that would intuitively have been considered as outliers.

7. Data quality

7.1 Overview of data quality evaluation

The objective of the 2012 APS is to produce quality estimates in the areas of education, employment, health and other core indicators for First Nations people living off reserve, Métis and Inuit aged 6 years and over.

Sections 7.2 and 7.3, below, explain the two types of errors that occur in surveys - sampling and non-sampling errors. Each type of error is evaluated in the context of the 2012 APS. Sampling error is the difference between the data obtained from the survey sample and the data that would have resulted from a complete census of the entire population taken under similar conditions. Thus, sampling error can be described as differences arising from sample-to-sample variability. Non-sampling errors are all other errors that are unrelated to sampling. Non-sampling errors can occur at any stage of the survey process, and include non-response for the survey as well as errors introduced during data collection or computer processing. Respondents may have made errors in their responses, trying to recall facts from the past, for example, or when a proxy stands in for a respondent. A response may have been incorrectly captured due to interviewer fatigue or a computer malfunctioning. An error may have been made in programming when the data were being processed or totalled. These are all examples of non-sampling errors.

This chapter describes the various measures adopted to prevent errors from occurring wherever possible and to adjust for any errors found throughout the different stages of the APS. Areas of caution for interpreting APS data are noted.

3. Bernier, J. et Nobrega, K. (1998). Outlier detection in asymmetric samples: A comparison of an inter-quartile range method and a variation of a sigma-gap method. Statistical Society of Canada Annual Meeting, June 1998.

7.2 Sampling errors and bootstrap method

The estimates that can be derived from the 2012 Aboriginal Peoples Survey are based on a sample of individuals. Somewhat different estimates might be obtained if a complete census had been taken using the same questionnaires, interviewers, supervisors, processing methods, etc. as those actually used. The difference between an estimate obtained from the sample and the one resulting from a complete count taken under similar conditions is called the "sampling error" of the estimate

In order to provide estimates of sampling error for statistics produced in the APS, a particular type of bootstrap method (the bootstrap being itself a specific resampling method), was developed. Several bootstrap methods exist in the literature but none of them was appropriate for the APS sampling design. The particularities of the APS design that made the estimation of sampling errors difficult were the following:

- Three-phase sampling design in which households (or dwellings) were selected in the first two phases and individuals in the third phase (section 3.2.3);
- The sampling fraction of the first phase sample (NHS sample) was non negligible (about 1/3 in the N1 regions) and the APS sampling fraction was generally relatively high in most strata;
- The APS strata (combinations of domains of estimation, N1 or N2 regions, initial respondent vs. NRFU respondent, identity vs. ancestry-only) were not nested within the NHS strata (collection units and groups of collection units);
- The method used had to be flexible enough to produce standard statistics such as proportions, totals, means and ratios but also more sophisticated statistics, including percentiles, logistic regression coefficients, etc.

Some of these particularities were encountered during the 2006 APS. However, in 2006, the survey frame was constructed from the Census long form. The sampling fraction was then about one in five households and the response rate was high because of the mandatory nature of the Census. In 2011, the NHS had an even larger sampling fraction of about one in three households. Since the NHS was a voluntary survey, non-response was higher than it had been for the 2006 long form questionnaire and a subsample of non respondents was selected for non response follow-up, which made the sampling design more complex. A more detailed description of the NHS sampling design is found in Chapter 3 of the *National Household Survey User Guide*.

Because non-response to the NHS was relatively important, **for the purpose of calculating variance only** (variance is a particular measure of sampling error), NRFU respondents were considered as a third phase sample, where the probability of inclusion of a household was equal to its own response probability and did not depend on the response probabilities of other households.

Several bootstrap methods exist in the literature for single-phase sampling and for multi-stage sampling. The most common one is called the "with-replacement bootstrap" and consists of selecting M with-replacement subsamples from the main sample and producing estimates for each subsample. The bootstrap variance estimate is then derived as a function of the squared differences between estimates coming from each of the M bootstrap subsamples and the estimate coming from the survey sample.

Variance calculation is greatly simplified though the use of bootstrap weights. For each subsample (bootstrap replication), the initial sampling weight first has to be adjusted for bootstrap subsampling, which produces what is called "initial bootstrap weights". Since each bootstrap sample is drawn by selecting the units with replacement, a unit can appear several times in a particular bootstrap sample. It can be shown that the bootstrap weights are a function of the initial sampling weight of the observation multiplied by what is called "the multiplicity" of the unit in the bootstrap sample, which is the number of times the unit is selected in the bootstrap sample. The multiplicity of a unit in the bootstrap sample is a random variable following what is called a "multinomial distribution". Hence, the bootstrap weights can be seen as the product of the initial sampling weights by a random adjustment factor (in this case, a function of the multiplicity of the unit). Once initial bootstrap weights have been derived, all weight adjustments applied on the initial sampling weights are applied to the initial bootstrap weights to obtain the final bootstrap weights, which will capture the variance associated with not only the particular sampling design but also the variance associated to all weight adjustments applied to the full sample to derive the final weights.

For the 2006 APS, a general bootstrap method for two-phase sampling[4] was developed.

As mentioned earlier, bootstrap weights can be seen as the product of the initial sampling weight by a random adjustment factor. This is the idea behind the general bootstrap methodology used in 2006. In the case of that two-phase sample, the variance was decomposed into two components, each one associated to a phase of sampling. The general two-phase bootstrap methodology produced a random adjustment factor for each phase of sampling. In the case of the 2006 APS, the initial bootstrap weight of a unit was the product of the initial sampling weight by these two random adjustment factors.

IIn 2011, however, **to estimate variance only**, the NHS was seen to have two additional phases, one corresponding to the NRFU subsample and one associated with non-response to the NRFU subsample. For the 2012 APS, the three phases of the NHS were combined into a single phase and the general two-phase bootstrap methodology (one NHS phase and one APS phase) was then applied. In the general bootstrap method for two-phase designs, random adjustment factors are functions of the simple and double inclusion probabilities associated with each of the phases. To combine the three phases into one, the simple and double inclusion probabilities were combined for the three NHS phases. The simple and double inclusion probabilities of the combined three phases are given by the product of the simple and double inclusion probabilities of each of the three phases. The details of the methodology used are found in Haddou (2013)[5].

Once the three NHS phases were combined into a single phase, the general bootstrap method for two-phase sampling was applied, which involved calculating two sets of random adjustment factors, that is one set for each phase.

The presence of these two sets of random adjustment factors had a major advantage. The first set could be used for estimates based on the first phase only, that is, estimates based on the NHS sample. These estimates were used when the weights were adjusted based on the NHS totals at the time of post-stratification (section 6.5). This produced variable NHS totals for each bootstrap sample and reflected the fact that NHS totals were based on a sample and not on known, fixed totals.

For the APS, 1000 sets of bootstrap weights were generated using the method described above. The method used is slightly biased upward in the sense that it slightly overestimates the variance. However, the amount of overestimation was found to be negligible for the APS. The method can also lead to negative bootstrap weights. To overcome this problem, a transformation was done on the bootstrap weights that reduced their variability. Therefore, the variance calculated on these transformed bootstrap weights has to be multiplied by a factor that is a function of a certain parameter, called *phi*. The value of the parameter corresponds to the smallest integer that makes all bootstrap weights positive. For the APS, this factor is 4. The variances calculated on the transformed bootstrap weights have to be multiplied by $4^2 = 16$. In addition, the CVs obtained (square root of the variance divided by the estimate itself) have to be multiplied by 4. However, most software which produce sampling error estimates from bootstrap weights have an option to specify this adjustment factor such that the correct variance estimate is obtained without the need of an extra step to multiply by the constant.

It is extremely important to use the appropriate multiplicative factor for any estimate of sampling error such as variance, standard error or CV. Omission of this factor would lead to erroneous results and conclusions. This factor is often specified as the "Fay adjustment factor" in software which produces sampling error estimates from bootstrap weights.

Note that if C is the variance multiplicative factor, some software use the parameter k instead where $k = 1\text{-}1/\sqrt{C}$. In our case, since C=16, then k=0.75. For examples of procedures using the Fay adjustment factor, see the *Aboriginal Peoples Survey, 2012: User's Guide to the Analytical File*.

4. Langlet, É., Beaumont, J.-F., and Lavallée, P. (2008). Bootstrap Methods for Two-Phase Sampling Applicable to Postcensal Surveys. Paper presented at Statistics Canada's Advisory Committee on Statistical Methods, May 2008, Ottawa.
5. Haddou, M. (2013). Bootstrap Variance Estimation Specifications - Aboriginal Peoples Survey. Internal document, January 2013.

The sampling error measure used for the APS is the coefficient of variation (CV) of the estimate, which is the standard error of the estimate divided by the estimate itself. When the CV of an estimate is less than or equal to 16.6%, the estimate can be used without restriction. In this survey, when the CV of an estimate is greater than 16.6% but smaller or equal to 33.3%, the estimate will be accompanied by the letter "E" to indicate that the data should be used with caution. When the CV of an estimate is greater than 33.3%, the cell estimate will be replaced by the letter "F" to indicate that the data is suppressed for reasons of reliability. An "X" is used to indicate that an estimate is suppressed to meet confidentiality requirements of the *Statistics Act*. These rules are summarised in Table 5.

Table 5 Sampling variability guidelines

Type of Estimate	c.v. (in %)	Guidelines	Symbol
Acceptable	c.v. ≤ 16.6	Estimates can be considered for general unrestricted release. Requires no special notation.	
Marginal	16.6 < c.v. ≤33.3	Estimates can be considered for general unrestricted release but should be accompanied by a warning cautioning subsequent users of the high sampling variability associated with the estimates. Such estimates should be identified by the letter E (or in some other similar fashion).	E – use with caution
Unacceptable	c.v. > 33.3	Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter F (or in some other fashion) and the following warning should accompany the estimates: "The user is advised that . . . (specify the data) . . . do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data."	F – too unreliable to be published

7.3 Non-sampling errors

Besides sampling, a number of factors at almost every stage of a survey can cause errors in survey results. Non-sampling errors arise primarily from the following sources: non-response, coverage, measurement and processing. For each of these areas, the following sections discuss the various measures used to minimize and correct error. For example, measurement errors may be due to respondents misunderstanding the questions and answering them inaccurately; also responses may be entered incorrectly during data capture and errors may be introduced in the processing and tabulation of data. Moving from a paper questionnaire in 2006 to Computer Assisted Interviewing (CAI) in 2012 greatly reduced the level of non-sampling error because CAI allows for the direct capture of responses, automated flows between questions, built in edits which eliminate inconsistencies and outliers, etc. (for more information on CAI, please refer to section 2.1).

Over a large number of observations, randomly occurring errors will have little effect on the estimates from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Thus, much time and effort was devoted to reduce non-sampling errors in the survey as described in the following sections.

7.3.1 Non-response errors

Non-response errors result from a failure to collect complete information on all units in the selected sample. Non-response produces errors in the survey estimates in two ways. First, non-respondents often have different characteristics from respondents, which can result in biased survey estimates if non-response is not corrected properly. The larger the non-response rate, the larger the risk of potential bias will be. Second, having a larger number of non-respondents reduces the effective size of the sample. As a result, the precision of the estimates decreases (the sampling error on the estimates will increase). This second aspect can be overcome by selecting a larger sample size initially. However, this will not reduce the potential bias in the estimates.

There are many types of non-response. One form of non-response is item non-response (or partial non-response), where the respondent does not respond to one or more questions, but has completed a significant portion of the overall questionnaire. Item non-response can be due to difficulty understanding a particular question.

Generally, the extent of item non-response was relatively small in the APS. Extensive qualitative reviews and testing of questionnaire was done prior to the survey, hence reducing the extent of item non-response. A response to key pre-defined questions was required before a case was classified as "respondent" as described in section 5.3.1. There were some cases, however, where a large proportion of responses to key questions were missing. These cases were eliminated from the database of respondents (did not satisfy definition of respondent) and were treated during weighting as a special case of total non-response (section 6.4). Finally, there is total non-response when the person selected to participate in the survey could not be contacted or did not participate once contacted. Weights of respondents were inflated in order to compensate for those who did not respond as described in section 6.3.

To mitigate the number of non-response cases, many initiatives were undertaken. In the months leading up to the survey, a comprehensive communications strategy was implemented to encourage participation as described in section 4. In addition, in-depth interviewer training was conducted. Interviewer training in conjunction with detailed interviewer manuals was done under by experienced Statistics Canada training staff, who oversaw activities in the field. Efforts to reach non-respondents through call-backs and follow-ups were also made by senior interviewers to encourage respondents to participate in the survey. Field follow-up, using CAPI interviewers, was also conducted in many specific regions.

A detailed table of final response rates obtained for the 2012 APS is provided in section 3.3 of this guide (Table 4).

7.3.2 Coverage errors

As mentioned in section 3.1, the target population of the 2012 APS was the Aboriginal identity population of Canada, aged 6 years and over as of February 1, 2012, living in private dwellings, excluding persons living on Indian reserves or settlements and in certain First Nations communities in Yukon and the NWT. The population sampled or covered by the survey corresponded to NHS respondents reporting Aboriginal ancestry or identity (see section 3.1.1) with the same restrictions as those for the target population in terms of age and geography. For data on First Nations people living on reserve, researchers are directed to use the 2011 NHS. Alternately, information on that population will be made available through the First Nations Regional Early Childhood, Education and Employment Survey conducted by the First Nations Information Governance Centre (see section 2.2).

Coverage errors occur when there are differences between the target population and the sampled population (population covered by the frame). Over-coverage is generally not an issue since out of scope units in the sample are typically identified during data collection and can be estimated for the entire survey frame. However, under-coverage can exist. Because the APS sample was selected from those who had participated in the NHS, individuals who did not participate in the NHS could not be sampled for the APS (the NHS had an unweighted response rate of 68.6%). As such, non-response bias in the NHS could translate to coverage bias in the APS (although technically, this could also be considered as a non-response bias for APS). Statistics Canada conducted several studies, before and after NHS data collection, to assess the risk and extent of the potential coverage bias in the NHS. A number of measures were taken to mitigate its effects. Namely, particular non-response follow-up procedures were used to reduce the potential bias for populations at risk such as the Aboriginal population. Particular weighting strategies were also used to reduce this bias. For a full discussion of data quality for the NHS, please refer to the *National Household Survey User Guide*.

7.3.3 Measurement errors

Measurement errors occur when a provided response differs from the real value. Such errors may be attributable to the respondent, the interviewer, the questionnaire, the collection method or the respondent's record-keeping system. Extensive efforts were made to develop questions for the 2012 APS which would be understood, relevant and culturally sensitive.

Following the release of data from the 2006 APS, an extensive content review was conducted of existing APS questions. The review brought together expertise from a diverse group of researchers and subject matter experts from within and outside of Statistics Canada. An analysis was conducted on which questions worked the best and which were most effective in producing valid indicators. This process also extended into an extensive search for relevant questions from other standardized survey questions at Statistics Canada.

Questions selected for potential inclusion on the 2012 questionnaire then underwent several rounds of qualitative testing using one-on-one interviews with respondents in eight different communities across various regions of Canada, including Iqaluit in the north. Testing was done among First Nations people, Métis and Inuit. Qualitative testing of the survey questionnaire was carried out by Statistics Canada's Questionnaire Design Resource Centre (QDRC). To minimize measurement error, adjustments were made to question wording and flows based on those results.

Many other measures were also taken to specifically reduce measurement error, including the use of skilled interviewers, extensive training of interviewers with respect to the survey procedures and content, and observation and monitoring of interviewers to detect problems of questionnaire design or misunderstanding of instructions.

7.3.4 Processing errors

Processing errors may occur at various stages of the survey process including data capture, coding and editing. Quality control procedures were applied to every stage of APS data processing to minimize this type of error. As compared to the 2006 APS, processing errors in 2012 were greatly reduced from the fact that the 2006 paper and pencil data collection method was replaced by Computer Assisted Interviewing.

APS questionnaires were first reviewed in the field by the interviewer supervisor. At the data processing stage, a detailed set of procedures and edit rules was used to identify and correct any inconsistencies between the responses provided. A set of thorough, systematized procedures was developed to assess the quality of every variable and to make corrections to any errors found. A snapshot of the output files was taken at each step and verification was done by comparing files at the current and previous step. The programming of all edit rules was exhaustively tested before being applied to the data. Some examples of the data processing verifications were: the review of all question flows, including very complex sequences, to ensure skip values were accurately assigned and distinguished from different types of missing values; quality control double-coding of "other-specify" responses; experienced supervision of coding to standardized classifications; and the review of all derived variables against their component variables to ensure accurate programming of derivation logic, including very complex derivations. See the data processing chapter (section 5) of this guide for more details.

8. Differences between the Aboriginal Peoples Survey and other data sources

Due to a number of differences in methodology between the 2012 Aboriginal Peoples Survey (APS), previous Aboriginal Peoples Surveys and other Statistics Canada surveys, comparisons of data between sources should be done with caution. The following sections of this document provide a detailed review of items affecting data comparability, providing data users with important information on factors to consider when conducting analyses with APS data.

8.1 Differences between the Aboriginal Peoples Survey and the National Household Survey

The APS selects its sample from respondents with specific responses in the NHS. More detailed information about how NHS responses were used to determine the population of interest for the APS is provided in section 3 (Survey design).

The NHS and the APS are both rich sources of information on Aboriginal peoples that complement each other. The APS takes concepts that are touched on in the NHS and asks questions that dig deeper in order to provide more detailed information. For instance, the NHS provides information on academic qualifications obtained (certificates, diplomas and degrees, from questions 27 to 30). Adding information from the APS provides an opportunity to learn more about any schooling below high school completion, whether teachers were Aboriginal people, whether financial assistance was obtained to pursue post-secondary schooling or why people did not continue their schooling.

The APS also covers entire topics or themes that are not included in the NHS. For example, the APS can provide detailed information on the employment and health of Aboriginal peoples.

Although both surveys cover the "identity population" by design, the 2012 APS, unlike the 2006 APS, did not cover the "ancestry-only population" described in section 3. (NHS respondents reporting Aboriginal ancestry-only were part of the APS sample because they had a non-negligible probability of reporting identity on the APS, and these respondents only remained in the APS data set if they actually reported Aboriginal identity in the APS).

In general, the Aboriginal identity population counts on the 2012 APS for certain subpopulations may differ from those obtained from the NHS, even if the population universe for the NHS is restricted to that of the APS. The second post-stratification described in section 6.5 ensured that the number of individuals with Aboriginal identity was the same in the NHS and the APS, but this applied only to certain combinations of Aboriginal group, region and age group. However, the Aboriginal identity population counts may differ for other subpopulations which were not controlled for during post-stratification. Moreover, for a given individual, the Aboriginal identity reported may differ in some cases between the NHS and the APS. There are a number of reasons why Aboriginal identity may not be the same on both surveys.

8.1.1 Different interview methods and impact of proxy reporting

In most regions, 2011 NHS data were collected by self-enumeration. Questionnaires were completed on the Internet or returned by mail. For Indian reserves and remote areas, including Inuit communities, Census interviewers were used. Often one member of the household completed the NHS form for all members of the household. This is called proxy reporting.

As described in section 4, the APS data were collected by computer-assisted interview (CAI) from the individual selected in the case of an adult and from the parent or guardian in the case of a child (direct interview was also allowed for those 15 to 17 years with the consent of the parent or guardian). Because the person contacted for the APS may not be the same person who filled in the Census questionnaire, there may be some differences in responses to similar questions.

8.1.2 Different questionnaires

The Aboriginal self-reporting question is one of the other sources of discrepancy between the NHS and the APS. For the 2012 APS, this question was divided into two parts as described in section 8.2.3. In addition, although there is no question on Aboriginal ancestry in the 2012 APS, the NHS asks a question on ethnic or cultural origins (question 17) immediately before the NHS question on Aboriginal self-reporting (question 18). Several Aboriginal origins are included as examples in the NHS ethnic or cultural origins question. This might affect the responses provided to question 18 on Aboriginal self-reporting.

The wording of the first question on Aboriginal self-reporting on the APS is slightly different from that of NHS question 18. Among other things, the term "North American Indian" in brackets after the term "First Nations" is not directly mentioned in the wording of the corresponding APS question. It is however explained in a note to the interviewer that the terms "First Nations" and "North American Indian" are interchangeable.

These differences in the questionnaire may lead to differences in how respondents answer.

8.1.3 Different contexts

The NHS questionnaire asks a limited number of general questions for the entire population of Canada, while the APS is specifically targeted to Aboriginal people. Consequently, given the more refined context of the APS, the concept of Aboriginal identity may be understood more clearly. Hence, it is possible that individuals who had reported being Aboriginal on the NHS may no longer report themselves as being Aboriginal in the APS. Conversely, individuals who had reported Aboriginal ancestry-only in the NHS may later report Aboriginal identity in the APS.

8.1.4 Effect of time

The concept of Aboriginal identity may not be a static characteristic in time. Events affecting the rights of certain Aboriginal groups or changes in the general population's perception of Aboriginal Peoples may affect the way in which Aboriginal identity questions are answered. Individuals who see themselves as having only Aboriginal ancestry at one point in time may later self-report as being Aboriginal. Furthermore, individuals may see themselves as First Nations people at a certain point in time and Métis at another.

8.1.5 Processing

It is very important to point out that, for the 2012 APS, all individuals who were either Status Indian or a member of a First Nation/Indian band but who had not also self-reported as Aboriginal were imputed as being a First Nations person on the APS. As such, this relatively small group in the NHS does not exist on the final APS data file. Hence, when comparing NHS and 2012 APS Aboriginal identity counts by Aboriginal group, NHS respondents who were either Status Indian or a member of a First Nation/Indian band and who did not self-report as Aboriginal should be included in the estimate for First Nations people.

8.1.6 Differences in the estimates

The following tables compare NHS estimates to APS estimates for different geographic regions and Aboriginal groups. The four Inuit regions are separated from the rest of Canada.

Table 6 compares NHS estimates to APS estimates for the Aboriginal identity population without double counting. This is the Aboriginal identity population aged 6 and over as of February 1, 2012, living in private dwellings excluding persons living on Indian reserves or settlements and excluding certain First Nations communities in Yukon and the NWT, which corresponds to the APS coverage. Note that for the NHS, there is a "Status Indian or member of a First Nation/Indian band only" category, which does not exist for the APS after imputation (cases such as these were imputed as First Nations people for the APS). A column has been created for the total of the NHS "First Nations" and "Status Indian or member of a First Nation / Indian band only" categories.

Table 7 is similar but reflects the double count. Hence, a person with a multiple identity of First Nations and Métis will be counted in the First Nations category as well as in the Métis category.

All counts in the next tables are rounded to the nearest 10. Since totals are rounded independently from individual cells, the cells may not add up exactly to the corresponding totals.

Table 6 Number of persons aged 6 and over[1] living off reserve[2] and reporting Aboriginal identity on the two surveys without double counting

	First Nations				Métis	
Region	NHS FN	NHS IB	NHS Tot	APS	NHS	APS
Nunatsiavut	10	0	10	20	20	10
Nunavik	50	0	50	50	40	10
Inuvialuit	790	10	800	730	180	210
Nunavut	110	10	130	100	130	90
Nunangat (total)	**960**	**20**	**980**	**890**	**360**	**320**
Atlantic excluding Nunatsiavut	36,650	4,050	40,700	40,760	21,540	21,330
Quebec excluding Nunavik	40,850	4,150	45,010	45,160	37,830	37,840
Ontario	140,250	7,220	147,460	149,030	80,060	79,980
Manitoba	45,510	900	46,420	46,910	71,450	71,470
Saskatchewan	40,680	900	41,580	41,640	46,420	46,360
Alberta	61,130	2,900	64,030	64,410	87,460	87,290
British Columbia	93,620	3,300	96,920	97,630	63,550	63,610
Yukon	4,050	40	4,090	4,070	710	710
Northwest Territories excluding Inuvialuit	3,250	120	3,370	3,340	1,880	1,870
Rest of Canada (total)	**465,990**	**23,580**	**489,570**	**492,960**	**410,900**	**410,460**
Canada (total)	**466,950**	**23,610**	**490,560**	**493,850**	**411,260**	**410,780**

Table 6 Number of persons aged 6 and over[1] living off reserve[2] and reporting Aboriginal identity on the two surveys without double counting (continued)

	Inuit		Multiple		Identity (total)	
Region	**NHS**	**APS**	**NHS**	**APS**	**NHS**	**APS**
Nunatsiavut	2,090	2,090	0	10	2,120	2,120
Nunavik	9,270	9,270	20	40	9,380	9,380
Inuvialuit	2,920	2,920	10	50	3,900	3,900
Nunavut	23,190	23,190	20	80	23,460	23,460
Nunangat (total)	**37,480**	**37,480**	**40**	**180**	**38,860**	**38,860**
Atlantic excluding Nunatsiavut	4,740	5,220	600	740	67,570	68,050
Quebec excluding Nunavik	1,520	1,130	1,360	1,200	85,720	85,340
Ontario	2,910	3,480	2,520	1,050	232,960	233,530
Manitoba	480	590	950	430	119,290	119,410
Saskatchewan	270	180	500	500	88,770	88,680
Alberta	1,780	1,340	1,590	1,370	154,860	154,420
British Columbia	1,460	1,320	2,100	1,330	164,030	163,890
Yukon	150	130	10	30	4,970	4,950
Northwest Territories excluding Inuvialuit	800	710	30	60	6,070	5,990
Rest of Canada (total)	**14,110**	**14,110**	**9,670**	**6,720**	**924,250**	**924,250**
Canada (total)	**51,580**	**51,580**	**9,710**	**6,900**	**963,110**	**963,110**

NHS National Household Survey
APS Aboriginal Peoples Survey
FN First Nations
IB Status Indian or member of a First Nation/Indian band only
Tot Total for First Nations or Status Indian or member of a First Nation/Indian band only
1. As of February 1, 2012.
2. And not living in certain First Nations communities in the Yukon and NWT.

Table 7 Number of persons aged 6 and over[1] living off reserve[2] and reporting Aboriginal identity on both surveys with double counting

	First Nations				Métis		Inuit	
Region	**NHS FN**	**NHS IB**	**NHS Tot**	**APS**	**NHS**	**APS**	**NHS**	**APS**
Nunatsiavut	10	0	10	20	20	10	2,090	2,100
Nunavik	60	0	70	80	40	20	9,290	9,320
Inuvialuit	800	10	800	760	180	220	2,930	2,960
Nunavut	120	10	130	170	140	100	23,210	23,270
Nunangat (total)	**990**	**20**	**1,010**	**1,030**	**380**	**360**	**37,520**	**37,650**
Atlantic excluding Nunatsiavut	37,090	4,050	41,140	41,140	22,050	22,030	5,000	5,630
Quebec excluding Nunavik	42,180	4,150	46,330	46,330	39,140	39,020	1,730	1,190
Ontario	142,670	7,220	149,890	149,890	82,440	80,940	3,230	3,750
Manitoba	46,430	900	47,330	47,330	72,390	71,900	520	610
Saskatchewan	41,170	900	42,070	42,070	46,920	46,810	270	290
Alberta	62,590	2,900	65,490	65,490	88,960	88,640	2,010	1,660
British Columbia	95,650	3,300	98,950	98,950	65,600	64,940	1,600	1,330
Yukon	4,060	40	4,100	4,100	730	740	160	140
Northwest Territories excluding Inuvialuit	3,270	120	3,390	3,390	1,900	1,920	810	740
Rest of Canada (total)	**475,110**	**23,580**	**498,700**	**498,700**	**420,120**	**416,940**	**15,330**	**15,330**
Canada (total)	**476,100**	**23,610**	**499,710**	**499,730**	**420,500**	**417,300**	**52,850**	**52,980**

NHS National Household Survey
APS Aboriginal Peoples Survey
FN First Nations
IB Status Indian or member of a First Nation/Indian band only
Tot Total for First Nations or Status Indian or member of a First Nation/Indian band only
1. As of February 1, 2012.
2. And not living in certain First Nations communities in the Yukon and NWT.

8.2 Differences between the 2006 APS and the 2012 APS

A number of major changes took place between the 2006 APS and the 2012 APS, not only in terms of survey content but also in terms of methodology. Because of these changes, caution should be exercised when comparing population estimates from the two surveys. In fact, the APS dissemination strategy has never been focused on the production of Aboriginal population counts. The strategy was, rather, to present the characteristics of the population for certain subgroups using proportions.

8.2.1 Methodological differences

In the case of the methodology, the most significant difference between the 2012 APS and the 2006 APS comes from the fact that the Aboriginal ancestry-only population was no longer part of the survey's target population in 2012 as it was in 2006. In fact, no question on Aboriginal ancestry was asked at the time of 2012 APS. It should be remembered that individuals reporting Aboriginal ancestry-only on the NHS were sampled because of their non-negligible probability of reporting identity on the APS. Since the 2012 target population no longer included the Aboriginal ancestry-only population, it was possible to adjust the weights at the time of the second post-stratification so that the identity counts matched those of the NHS for certain combinations of region, Aboriginal group and age group (section 6.5). In 2006, however, it was not possible to adjust both the identity and ancestry-only counts. The second post-stratification had ensured that the total Aboriginal population (identity or ancestry), as estimated from the APS screening questions, matched those from the Census screening questions. This post-stratification had been carried out by geographic region and age group, based on the total number of Aboriginal people on the 2006 Census and not by each Aboriginal group.

These two very distinct post-stratification strategies between 2006 and 2012 mean that growth in the Aboriginal population cannot be measured from the 2006 and 2012 APS. In 2006, because there were fewer shifts from Aboriginal identity to Aboriginal ancestry-only or from Aboriginal identity to non-Aboriginal (neither identity nor ancestry) than there were shifts between Aboriginal ancestry-only and Aboriginal identity between the Census and the APS, the Aboriginal identity counts according to the 2006 APS were higher than those of the 2006 Census and the Aboriginal ancestry-only counts were lower on the 2006 APS than on the 2006 Census.

Another important difference in methodology is the fact that the 2006 APS sample was selected from respondents to the 2006 Census, while the 2012 APS sample was selected from respondents to the 2011 NHS. The characteristics of respondents to the NHS may be different than those of respondents to a census. The fact that non-respondents have different characteristics than respondents creates what is called non-response bias. Despite the fact that the NHS used follow-up strategies and non-response adjustment strategies at weighting to reduce this bias, it is possible that some non-response bias still remains. See section 7.3.2 for more details.

8.2.2 Content differences

As for content, section 8.1.2 explains that one major difference between the two most recent cycles of the APS was the fact that the question on Aboriginal self-reporting was divided into two parts on the 2012 APS (details are given in the following section). This could clearly lead to differences in how individuals respond in terms of Aboriginal self-reporting. Another very important difference is the fact that, in 2012, the Aboriginal self-reporting question was no longer preceded, as was the case in 2006, by three questions on Aboriginal ancestry (three questions in one). These three 2006 questions on Aboriginal ancestry probably meant that individuals were more likely to self-report as Aboriginal to the next question and were more likely to report multiple identities. Indeed, the proportion of individuals with multiple Aboriginal identities was much higher in 2006 than in 2012.

8.2.3 Changes to Aboriginal identity questions

Several changes were made to the 2012 APS questions measuring Aboriginal identity compared to the corresponding questions in the 2006 APS. Some of these changes reflect those introduced on the 2011 National Household Survey compared to the 2006 Census. Changes were made in order to enhance measurement accuracy, to reflect current terminology and to reflect recent legislative changes.

Self-reported Aboriginal group

For the 2012 APS, the measurement of self-reported Aboriginal group was divided into two parts, as shown in the table below. This was intended to enhance measurement by allowing respondents to self-report as Aboriginal in question ID_Q01 even if they did not identify with a specific group as named in ID_Q02. For 2012, the term "First Nations" is introduced for the first time. In 2006, the term "North American Indian" was used solely. In addition, the 2012 question clarified to respondents that the term "First Nations" includes both Status and Non-Status Indians. Finally, 2012 included both the terms "Inuk" and "Inuit", whereas 2006 only referred to "Inuk".

Table 8 Self-reported Aboriginal group, 2012 APS, 2006 APS, 2011 NHS and 2006 Census

2012 APS	2006 APS
ID_Q01 - Are you an Aboriginal person, that is, First Nations, Métis or Inuk (Inuit)? First Nations includes Status and Non-Status Indians. ID_Q02 - Are you First Nations, Métis or Inuk (Inuit)? First Nations (North American Indian) Métis Inuk (Inuit)	Are you an Aboriginal person, that is, North American Indian, Métis or Inuk?
2011 NHS	**2006 Census**
Is this person an Aboriginal person, that is, First Nations (North American Indian), Métis or Inuk (Inuit)?	Is this person an Aboriginal person, that is, North American Indian, Métis or Inuit (Eskimo)?

Note: First Nations (North American Indian) includes Status and Non-Status Indians.

Status Indian (Registered or Treaty Indian)

For the 2012 APS, the main question measuring "Status Indian" (ID_Q03) was re-worded with the term "Status" being used for the first time. Subsequent questions were updated to reflect a change to legislation in 2011: an amendment to the *Indian Act* called Bill C-3, *Gender Equity in Indian Registration Act*.

Table 9 Status Indian, 2012 APS, 2006 APS, 2011 NHS and 2006 Census

2012 APS	2006 APS
ID_Q03 - Are you a Status Indian, that is, a Registered or Treaty Indian as defined by the *Indian Act* of Canada?	Are you a Treaty Indian or a Registered Indian as defined by the *Indian Act* of Canada?
ID_Q04A - Have you ever applied to Aboriginal Affairs and Northern Development Canada (previously named INAC) to be registered as a Status Indian under Bill C-31 or Bill C-3?	Have you ever applied to the Department of Indian Affairs and Northern Development to be registered as a status Indian under Bill C-31?
ID_Q04B - Have you been registered as a Status Indian under Bill C-31 or Bill C-3?	Have you been registered as a Status Indian under Bill C-31?
2011 NHS	**2006 Census**
Is this person a Status Indian (Registered or Treaty Indian as defined by the *Indian Act* of Canada)? (asked before First Nation/Indian band question for first time)	Is this person a Treaty Indian or a Registered Indian as defined by the *Indian Act* of Canada?

Member of a First Nation or Indian band

For the 2012 APS, the term "First Nation" precedes the term "Indian band"; in 2006 the term "Indian band" came first.

Table 10 Member of a First Nation or Indian band, 2012 APS, 2006 APS, 2011 NHS and 2006 Census

2012 APS	2006 APS
ID_Q05 - Are you a member of a First Nation or Indian band?	Are you a member of an Indian Band or First Nation?
2011 NHS	**2006 Census**
Is this person a member of a First Nation/Indian band?	Is this person a member of an Indian Band / First Nation?

9. Data dissemination

9.1 An overview of 2012 Aboriginal Peoples Survey dissemination

Data for the 2012 Aboriginal Peoples Survey (APS) were released publically on November 25, 2013. The APS release included an analytical article and short video, both available to the public free of charge on Statistics Canada's website. A set of regional tables designed to accompany the analytical article has been made available on-demand via e-mail. Additional custom data tables are also available on a cost-recovery basis.

Later releases of APS data, mostly in 2014, will include data tables, other analytical papers, a public use microdata file (PUMF) and an APS analytical file. Each of these products is discussed in more detail below.

Statistics Canada also delivers special APS presentations to National Aboriginal Organizations, to researchers working out of Research Data Centres, to other key stakeholders and at various conferences. In addition, Statistics Canada's network of Aboriginal Liaison Program Advisors across Canada will be distributing APS-focused newsletters and responding to APS data needs in their region.

9.2 Data products and services

A set of data tables on the themes of education and employment will be available shortly after the first 2012 Aboriginal Peoples Survey release on November 25, 2013. These tables compliment the release article (see section 9.3 below for article description), providing data at the national and regional levels for First Nations living off-reserve, Inuit and Métis and comparing high school leavers and completers between the ages of 18 and 44.

In 2014, a more extensive set of APS indicators will be available on CANSIM, the Canadian Socioeconomic Information Management system at Statistics Canada. The 2012 APS is the first APS for which CANSIM tables will be available. CANSIM provides fast and easy access to a large range of the latest statistics available in Canada. CANSIM tables are available on-line to the public free of charge.

An analytical file for the 2012 APS will be available in Statistics Canada's Research Data Centres (RDCs) across the country. In order to access the file, researchers must undergo a research and ethics committee review for approval. Their use of the data must be conducted according to Statistics Canada policies, guidelines and standards (for instance, only aggregate statistical estimates that conform to the confidentiality provisions of the *Statistics Act* may be released outside of Statistics Canada). For more information on the Research Data Centres Program, please refer to the Frequently Asked Questions.

Accompanying the analytical file is the record layout, SAS (Statistical Analysis System) and SPSS (Statistical Package for the Social Sciences) syntax to load the file, as well as metadata in the form of a codebook that describes each variable and provides weighted and unweighted frequency counts. In addition, a detailed user guide provides researchers with guidelines they need for conducting statistical analysis of 2012 APS data.

A public use microdata file (PUMF) will also be disseminated in 2014, allowing for wider and direct use of the data by researchers, including university students across the country. In order to provide extra assurance with respect to the non-disclosure of confidential information, the level of detail of the PUMF is not as fine as that of the analytical file kept by Statistics Canada.

The 2012 APS also has four data sharing agreements in place with each of the Inuit regions of Canada: Nunatsiavut (Northern coastal Labrador), Nunavik (Northern Quebec), the territory of Nunavut and the Inuvialuit region of the Northwest Territories. These data sharing agreements are an excellent method to increase data use and make the data more accessible. Only those respondents who are Inuit and have agreed to share their information with their respective region will be part of the data sharing file. The data sharing files will be disseminated in accordance with the data sharing agreements in 2014.

Another means of access to the data file is the Real Time Remote Access (RTRA) tool at Statistics Canada. This is a subscription service provided for a fee to clients associated with an academic institution, a government department or a non-profit organization. RTRA is an on-line remote access tool allowing users to run SAS software programs, in real-time, against micro-data sets located in a central and secure location. Researchers using the RTRA system do not gain direct access to the micro-data and cannot view the content of the micro-data file. Instead, users submit SAS programs to extract results in the form of aggregated frequency tables. The RTRA complements existing methods of access to confidential micro-data, including the Research Data Centres. Using a secure username and password, the RTRA provides around the clock access to survey results from any computer with internet access. Confidentiality rules and reliability guidelines are applied to all requests in an automated way by the RTRA system, eliminating the need for manual intervention and allowing for rapid access to results. An RTRA agent can be reached at: rtra_adtr@statcan.gc.ca or please visit www.statcan.gc.ca/rtra for more information on how to obtain a user account.

In addition to these data products and services, clients can request custom data tables from Statistics Canada Client Services by email at sasd-dssea@statcan.gc.ca or by phone at 1-800-263-1136. All custom requests are screened for confidentiality and aggregate data are rounded before being released to clients.

9.3 Analytical products

At the heart of the first 2012 Aboriginal Peoples Survey release is an analytical article entitled: *The Education and Employment Experiences of First Nations People Living Off Reserve, Inuit, and Métis: Selected Findings from the 2012 Aboriginal Peoples Survey*. In the article, an in-depth analysis of education and employment experiences is provided separately for First Nations people living off reserve, Inuit and Métis between the ages of 18 and 44. For each group, the article compares high school completers and leavers on key indicators of student experiences, the school environment and family involvement. The article also examines post-secondary education and employment experiences among high school completers and leavers. The results of this article were also released in the form of a summary video.

Other analytical papers, which provide highlights of 2012 APS data on specialized topics, are to be released in 2014. These will focus on the following themes:

- Inuit health
- Métis employment
- Off-reserve First Nations youth education

9.4 Reference products

Information about the Aboriginal Peoples Survey is available on Statistics Canada's website. Statistics Canada provides an Integrated Metadata Base (IMDB) on-line for all surveys that it conducts, including the 2012 Aboriginal Peoples Survey. The purpose of the IMDB is to provide information that will assist the public in interpreting Statistics Canada's published data. The information (also known as metadata) is provided to ensure an understanding of the basic concepts that define the data, including variables and classifications; the underlying statistical methods and surveys; and key aspects of the data quality. Direct access to the 2012 APS questionnaire is also provided.

As well, this Concepts and Methods Guide is provided online for a detailed discussion of survey content, sampling design, data collection and processing, weighting of the data, data quality, differences between the 2012 APS and 2011 National Household Survey, and dissemination products for the APS.

For researchers using the Analytical File in Statistics Canada's Research Data Centres (RDCs), the *Aboriginal Peoples Survey, 2012: User's Guide to the Analytical File* is available which details the concepts and methods of the survey along with detailed step-by-step instructions for using the data file. The RDC User Guide describes the structure of the data file in detail, including all core variables, derived variables and linkages to the National Household Survey (NHS). A detailed codebook provides the data dictionary for all variables available. The *Aboriginal Peoples Survey, 2012: User's Guide to the Analytical File* also provides detailed guidelines for tabulation and statistical analysis, how to apply the necessary weights to the data, information of software packages available and guidelines for the release of data, such as rounding rules. The process of estimating the reliability of estimates, both quantitative and qualitative, is covered in detail.

Finally, a separate *PUMF User Guide* is created for users of the Public use microdata file for APS. This covers similar issues of the variables available, a data dictionary codebook, the process of estimation, use of weights and guidelines for tabulation, statistical analysis and dissemination of data.

9.5 Disclosure control

Statistics Canada is prohibited by law from releasing any information it collects which could identify any person, business, or organization, unless consent has been given by the respondent or as permitted by the *Statistics Act*. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data.

Appendix A 2012 Aboriginal Peoples Survey indicators by questionnaire module

The list below includes questionnaire indicators as well as derived variables (DVs) which combine different questionnaire items.

Please note that this is not a complete list of all variables available from the 2012 Aboriginal Peoples Survey (APS): for a complete list of variables, please refer to the 2012 APS Codebook (data dictionary). For details on how to obtain the codebook, please contact Statistics Canada Client Services at sasd-dssea@statcan.gc.ca or 1-800-263-1136.

Introduction

- Purpose of the survey
- Voluntary nature of survey
- Explanation of data linkage with NHS

Aboriginal identification

- Aboriginal Group – First Nations, Métis, or Inuit
- Status – Registered or Treaty Indian as defined by the *Indian Act* of Canada
- Application or registration as Status Indian under Bill C-31 or C-3
- Member of a First Nation or Indian band

Aboriginal identity DVs include single and multiple identity groups and replications of historical definitions for comparability to previous surveys. Aboriginal ancestry is taken from the NHS since it is not measured directly by the 2012 APS

Household composition and marital status

- Age, sex, marital status of "Person Most Knowledgeable" responding for child and their relationship to the child
- Number of household members, and number under 18
- Family members living in household

Household composition DVs include age groups, number of people in the household, number of children, presence of partner, and number of generations in the household

Mobility

- Lived in current community entire life
- Reasons for moving to current community
- Number of years since moving to current community
- Ever lived on reserve or in an Inuit community
- Ever lived outside an Inuit community
- Reasons for moving from community (Reserve, Inuit community only)
- Reasons for remaining in community
- Frequency of moves

Mobility DVs include residence in a First Nations or Inuit community, number of years since left or returned from a First Nation or Inuit community

Education

Education status

- Current school enrolment
- Current level of schooling
- Highest level of schooling (derived based on info on post-secondary)
- High school (and high school equivalency) completion
- High school (and high school equivalency) attendance
- Location of high school equivalency program

Education streams

Education 1, 2, 3: grades 1-6, grades 7-12, completers-leavers under age 45
(Note that most but not all of these indicators were measured for all three groups.)

- Preschool attendance (including early childhood development program)
- Preschool designed for Aboriginal children (including early childhood development program)
- Number of schools attended
- Reason for changing schools
- School in First Nations community (on-reserve)
- Mobility for the purpose of attending school
- Exposure to Aboriginal languages: Teaching an Aboriginal language and teaching subjects in an Aboriginal language (other than language courses)
- Academic performance
- Support – tutoring: need, reasons, frequency, source
- Communication with parent/guardian
- Perceptions of school environment
- Family involvement with school
- Support of learning at home
- Frequency of reading by student
- Access to Internet
- School absences/skips/late arrivals: Frequency and reasons
- Expectations/aspirations for education
- Planning and savings for education
- Family education history: sibling, mother, father

Education 2, 3: Grades 7-12, Completers-leavers under age 45 - additional indicators

- Need for social support and source of support
- Sport or physical activities: Participation, frequency, within school, outside of school or both
- Art/drama/music: Participation, frequency, within school, outside of school or both
- Club or group activities: Participation, frequency, within school, outside of school or both
- First Nations/Métis/Inuit activities: Participation, frequency
- Time with Elders
- Volunteering/work without pay
- Work in the community (babysitting, at a store, tutor)
- Attitudes and behaviour of peers (in high school)
- Dropping out of school: Frequency and reasons

Education 3 & 4: High school leavers and completers, under age 45 and 45 and over

(Note that most but not all of these indicators were measured for both groups.)

- Reasons for returning to school
- Age of completing high school
- Age last attended elementary or high school
- Education towards postsecondary certification
- Type of educational institution, current
- Certificate(s)/Diploma(s)/Degree(s) completed
- Major field of study
- Current attendance at school (postsecondary)
- Postsecondary completion
- Year left/finished school
- Reasons for leaving postsecondary education
- Full-time or part-time studies
- Current field of study
- Program for preparing students
- Move to attend postsecondary education
- Distance education
- Personal support
- Government student loan
- Method of financing
- Financial barriers to postsecondary education
- Reasons for not pursuing a postsecondary education
- Plans for future education
- Family education history: sibling, mother, father

Education DVs include over 300 DVs combining equivalent questions across educational streams and 50 other question combinations such as highest level of schooling, major field of study, family involvement with school, need and receipt of support

Aboriginal languages

- Aboriginal language(s) spoken/understood
- Primary Aboriginal language
- Self-rated ability to speak and understand an Aboriginal language
- Importance of speaking and understanding an Aboriginal language
- Exposure to an Aboriginal language
- Language first learned in childhood

Aboriginal language DVs include over 50 DVs reflecting the complexity of measuring single and multiple languages spoken and understood

Residential school

- Attendance at residential school
- Family members' attendance at residential school

Residential school DVs include residential school attendance by respondent and family members

Employment (aged 15 and over)

Labour market activities

- Rates of employment, unemployment and participation
- Reason(s) for being absent from work

Labour force status

- Looking for work
- Reasons for not seeking employment
- Average hours sought
- Able to work
- Reason for not being available to work

Looking for work

- Methods for looking for work
- Duration of unemployment
- Barriers in finding work

Discouraged workers

- Seeking employment
- Reasons for not seeking employment
- Able to work
- Reason for not being available to work

Other employment modules

- Past job attachment
- Multiple employment
- Class of worker
- Incorporated business
- Job tenure
- Industry
- Occupation
- Usual hours of work

Part-time employment

- Main reason for working part-time

Permanent work

- Permanent work

Labour mobility

- Labour mobility
- Reasons for not moving to find employment elsewhere

Employment DVs include labour force status, full-time or part-time employment status, permanent and non-permanent work status, multiple job holder status

Traditional activities

- Making clothing or footwear
- Making arts or crafts for example carvings, drawings, jewellery
- Hunting, fishing or trapping
- Gathering wild plants for example berries, rice or sweet grass
- Participation in these traditional activities and frequency
- Reason for participating in traditional activities
- Barriers to participation
- Interest in participating

Traditional activities DVs include participation in a traditional activity in past year, participation for income, and interest in traditional activities

Income

- Sources and main source of personal income
- Total personal income
- Employment income

Income DVs include source of personal income in 2011, total personal income groups, employment income groups

Health

General health 1 – Health status

- Self-rated health status

Pregnancy and childbirth

- Pregnancy
- Number of births
- Age at time of first child

Height and weight

- Height and weight

General health 2 – Contact with health professionals

- Contact with regular medical doctor
- Barriers to contact with regular medical doctor
- Contact with various health professionals (e.g. dentist, psychologist)
- Barriers to contact with various health professionals
- Type of care needed

Chronic conditions

- Chronic conditions – under 12 years of age
- Physical limitations (children with chronic conditions)
- Chronic conditions – aged 12 and over

Injuries

- Injuries
- Frequency of injuries
- Type of injury
- Location when injured
- Cause of injury
- Injury(s) as a result of a fall and cause

Mental health

- Self-rated mental health

Distress

- Distress scale

Suicide

- Ideation of suicide
- Attempts at suicide

Smoking

- Smoking frequency
- Age began smoking
- Exposure to second-hand smoke in the home

Alcohol use

- Frequency of alcohol consumption
- Frequency of drinking 5 or more drinks on one occasion

Drug use

- Street drug use
- Prescription drug use

Food security

- Food security scale

Community support

- Community support

Health DVs include height, weight, body mass index, distress scale, suicide and support

Housing

- House owned or rented
- Dwelling in need of repairs
- Crowding, number of rooms
- Subsidized housing

Housing DVs include persons per room and length of time waiting for subsidized housing

Geography

Geography DVs include multiple NHS geographies and Inuit regions

Appendix B Extra response categories created for "Other – specify" questions

The table below summarizes the extra categories added to certain survey questions during the coding of data from questions involving an "Other – specify" category.

Theme	Indicator	Questions	Extra categories
Residential School	Family members' attendance at residential school	RS_S02F	RS_Q02G - Any aunts or uncles RS_Q02H - Any cousins
Health	Barriers to contact with regular medical doctor	GH2_S02	07 - Doctors always changing in the area; no permanent doctors available
Health	Type of care needed	GH2_S08	06 - Dental care
Employment	Reasons for not moving to find employment elsewhere	LM_S02	15 - Health problem or disability 16 - Happy / Comfortable here
Mobility	Reasons for remaining in community	MOB_S13	14 - Culture / traditional way of life
Traditional Activities	Reasons for participating in traditional activities: making clothes and making crafts	TA_S01C TA_S02C	05 - For school project 06 - For gifts, charity or fundraising
Traditional Activities	Barriers to participation (gathering)	TA_S04E TA_S04G	07 - Physical disability, health or medical reasons
Education	Location of high school equivalency program	EDS1_S06 EDS2_S05 ED4_S04	5 - At a college or technical institute
Education	Reasons for returning to school	ED3F_S39	09 - For my children / To be a role model
Education	School absences/skips/late arrivals Reasons	ED1_S28B ED2F_S37B	16 - Sleeping-in / Tired 17 - To participate in sports or cultural activities
Education	Reasons for no support-tutoring	ED2B_S17 ED3B_S16	07 - Student not interested, unwilling 08 - Looking into it, application in process 09 - Tutoring not sought out or pursued 10 - Student dropped out of school / Attendance problems

Appendix C Standard classifications used for the 2012 Aboriginal Peoples Survey

Major field of study: Classification of Instructional Programs

Respondents of the 2012 APS provided information on the major field of study of the highest certificate, diploma or degree that they completed. Responses were coded according to the Classification of Instructional Programs (CIP) Canada 2011.

The CIP was originally developed in 1980 by the National Center for Educational Statistics (NCES) in the United States. NCES released updates in 1985, 1990, 2000 and 2010. CIP Canada 2011 is the second Canadian version of this classification; the first Canadian version being CIP Canada 2000.

CIP is designed to classify 'instructional programs'. The organizing principle behind CIP is 'field of study'. At Statistics Canada, a field of study is defined as a "discipline or area of learning or training".

Industry and occupation:North American Industry Classification System and National Occupational Classification

APS respondents provided information on their occupation and the industry in which they worked during the reference week of the survey. These responses were coded using the North American Industry Classification System (NAICS) Canada and the National Occupational Classification (NOC). Data were coded to the 4-digit level of detail and were coded to both current/emerging and historical versions of these classification systems, as follows:

NAICS Canada 2012 - North American Industry Classification System (NAICS) Canada

NAICS Canada 2007 - North American Industry Classification System (NAICS) Canada

NOC – 2011: National Occupational Classification (NOC)

NOC-S – 2006: National Occupational Classification - Statistics (NOC-S)

NAICS is an industry classification system developed by the statistical agencies of Canada, Mexico and the United States. It was created against the background of the North American Free Trade Agreement. The 2012 NAICS revision was undertaken to achieve one main goal: to modify or create industries to reflect new, emerging, or changing activities and technologies. NAICS is based on supply-side or production-oriented principles - the criteria used to group establishments into industries in NAICS are similarity of input structures, labour skills and production processes.

NOC 2011 is the nationally accepted taxonomy and organizational framework of occupations in the Canadian labour market. It updates both the National Occupational Classification of Human Resources and Skills Development Canada and Statistics Canada's National Occupational Classification for Statistics (NOC-S). As in the past, this revision is based on extensive occupational research, analysis and consultation conducted across the country. Occupations are identified and grouped primarily in terms of the work usually performed, this being determined by the tasks, duties, and responsibilities of the occupation.

Aboriginal languages

For the coding of Aboriginal languages captured on the 2012 APS, 51 Aboriginal language categories were used. Coding was conducted in accordance with the 2011 Census and the 2011 National Household Survey Classification. This differs from the classifications used in 2006 and 2001. For instance, the Michif language family has a new separate category in 2011 and Inuit languages are now captured in greater detail. For details on the complete classification and the concordances with previous years, see the *2011 Census Dictionary*, Appendix D.

Appendix D Glossary of survey terms

A

Analytical file

A Statistics Canada **microdata** set for a given survey, available for use in Research Data Centres (RDCs) across Canada. RDCs provide researchers with access, in a secure university setting, to microdata from population and household surveys. The centres are staffed by Statistics Canada employees. They are operated under the provisions of the ***Statistics Act*** in accordance with all the confidentiality rules and are accessible only to researchers with approved projects who have been sworn in under the *Statistics Act* as 'deemed employees.'

Aboriginal peoples

Aboriginal peoples in Canada include **First Nations people**, **Métis** and **Inuit**.

B

Bootstrap method

The bootstrap method is an approach for estimating error in a dataset related to **sampling**. Sampling introduces error because data are not taken from the entire population, but only a sub-section, called a sample, which is then used to make **estimates** for the whole population. There are several methods for estimating the level of **sampling error**. The bootstrap method usually selects a number of subsamples from the main sample and produces estimates for each subsample. The sampling error is estimated as a function of the observed differences between estimates from the different subsamples and estimates from the complete sample.

C

Census metropolitan area (CMA) and Census agglomeration (CA)

Area consisting of one or more neighbouring municipalities situated around a major urban core. A census metropolitan area must have a total population of at least 100,000 of which 50,000 or more live in the urban core. A census agglomeration must have an urban core population of at least 10,000.

Census subdivision (CSD)

This is the general term for municipalities (as determined by provincial/territorial legislation) or areas treated a municipal equivalents for statistical purposes (e.g., Indian reserves, Indian settlements and unorganized territories).

Census of population

A census is the collection of information about all units in a population, sometimes also called a 100% sample survey. Under the ***Statistics Act*** of 1971, it is a statutory requirement to conduct a nationwide census every five years. The Census of Population provides information needed to plan community services such as schools, day care, police services and fire protection, to forecast consumer demand and to conduct market research studies.

Cohort

As used in demography, a number of people having a common characteristic, for example, all persons in a given population who were born in 1940, or all persons suffering from a particular disease.

Confidential information

This is a term used within Statistics Canada to describe information that is subject to the secrecy provisions of the ***Statistics Act***. Information is deemed confidential either because it directly identifies a **responding unit**, for example, by name, or because it could permit specific responding units to be identified, even when the data is stripped of identifiers, due to the information's detail or its geographical structure or format.

Confidentiality

Confidentiality denotes an implied trust relationship between the person providing the information and the individual or organization collecting it. This relationship is built on the assurance that the information will not be disclosed without the person's permission. Under the ***Statistics Act***, information that would identify an individual, business or institution cannot be disclosed without their knowledge or consent.

Coverage

Coverage is the extent to which every person or unit intended for inclusion in a survey or census is in fact counted and counted only once. Coverage errors refer to when persons or units of the survey or census are missed (under-coverage) or over-counted (over-coverage). Studies are often conducted by Statistics Canada to provide estimates of under-coverage and over-coverage of a given survey or census or to examine related issues. For example, Statistics Canada has studied and analyzed the extent to which cell-phone use affects coverage for telephone surveys.

CV – Coefficient of variation

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. In this process of estimation, some level of error is inevitable. The coefficient of variation (CV) is a way of expressing the **sampling error** associated with an **estimate**. First a **standard error** or 'average' error of the estimate is calculated. The CV is obtained by dividing the standard error of the estimate by the estimate itself and expressing the resulting fraction as a percentage. The lower the CV, the higher the data quality (see **margin of error**).

D

Data

Observations and measurements collected during a survey, census or other study. Facts or figures from which conclusions can be drawn.

Data quality

A degree or level of confidence that the data and statistical information are "fit for use". The particular issues of quality or fitness for use that must be addressed by Statistics Canada are relevance, accuracy, timeliness, accessibility, **interpretability** and coherence.

Dataset, database

An organized and sorted list of facts or information about a set of individuals, households, businesses, or other relevant units. A Statistics Canada dataset is usually generated by a survey or administrative data, stored on a computer, and organized in such a way that it may be accessed easily by a wide variety of statistical application programs.

Dissemination

The process of providing statistical products and services to the general public and to specific data users. Statistics Canada disseminates data and analysis in the form of survey results, research reports, technical papers, periodical magazines, census products, and research compendia. Online products date from 1996 to the present. Historical material can be located using the Library Catalogue. Statistics Canada information is also distributed to an approved network of depository libraries.The objective of dissemination activities is to provide relevant information in a timely fashion, in useful formats, and through accessible channels. Activities in place to support the dissemination of products include client consultation services, marketing, promotions, user-training and other client services.

Derived variable

A new **variable** constructed by applying logical or mathematical operations to one or more existing variables in order to meet particular data needs. For example, an age variable can be derived from date of birth information. As another example, a derived variable could be obtained called 'presence of a chronic health condition' based on whether or not a respondent answered 'yes' at least once to a series of questions asking about specific chronic health conditions such as asthma, diabetes, heart disease, etc.

E

Editing

Editing is a process that ensures survey data are accurate, complete and consistent. A set of editing rules or conditions is applied to a **dataset**. Data which do not meet the conditions are examined and corrected where appropriate.

Errors

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. The accuracy of such an **estimate** is a measure of how much the estimate differs from the correct or "true" figure. Departures from true figures are known as errors. Errors can arise from many sources, but can be grouped into a few broad categories: coverage errors, non-response errors, response errors, processing errors and sampling errors.

Coverage errors

Coverage errors refer to when persons or units of the survey are missed (under-coverage) or overcounted (over-coverage).

Non-response errors

Non-response errors occur when it proves impossible to obtain a complete questionnaire from a person, household, or organization. Although certain adjustments for missing data can be made during processing, non-response means that some loss of accuracy is inevitable.

Response errors

Response errors indicate that a response may not be entirely accurate. The respondent may have misinterpreted the question or may not know the answer, especially if it is given for an absent household member, for example.

Processing errors

Processing errors include mistakes made during data entry, coding, tabulation or other forms of data manipulation.

Sampling error

Sampling error refers to the fact that the results of the weighted sample differ somewhat from the results that would have been obtained from the total population. The difference is known as sampling error. The actual sampling error is of course unknown, but it is possible to calculate an "average" value, known as the "**standard error**".

Estimation, estimate

Using results of the weighted sample to estimate the characteristics of the total population.

F

First Nations, First Nations people

A term that came into common usage in the 1970s to replace the word "Indian," which many people found offensive. Although the term First Nations is widely used, no legal definition of it exists. Among its uses, the term "First Nations people" refers to the **North American Indian** people in Canada, both Status and Non-Status. Many people have also adopted the term "First Nation" to replace the word "band" in the name of their community.

Frame

A list, map, or conceptual specification of the units comprising the survey population from which persons can be selected. For example, a telephone or city directory, or a list of members of a particular association or group.

Frequency

The number of times an event or item occurs in a **dataset**.

Frequency distribution

A chart or table showing how often each value or range of values of a **variable** appear in a **dataset**. It is sometimes called a one-way frequency table to indicate that the distribution contains counts for one variable only.

G - H - I

Imputation

Imputation involves replacing either missing or invalid data with valid data. This is normally performed using predetermined rules or with the use of data from a 'statistical neighbour'–another **responding unit** who has similar characteristics. Imputation is often combined with data **editing**.

Indian Act

The Canadian federal legislation, first passed in 1876, that sets out certain federal government obligations, and regulates the management of Indian reserve lands. The act has been amended several times, most recently in 1985.

Indian band

A group of **North American Indian** people for whom lands have been set apart and money is held by the Crown. Each band has its own governing band council, usually consisting of one or more chiefs, and several councillors. Community members choose the chief and councillors by election, or sometimes through traditional custom. The members of a band generally share common values, traditions and practices rooted in their ancestral heritage. Today, many bands prefer to be known as **First Nations**.

Information

Data that have been recorded, classified, organized, related or interpreted within a framework so that meaning emerges.

Information product

Organization of results from Statistics Canada activities, including data files, **databases**, tables, graphs, maps, and text. This organization can be either pre-defined (standard information product) or made in response to special requests (customized information product). Information products can be made available on either print or electronic media.

Interpretability

Interpretability reflects the ease with which the user may understand, properly use and analyze the data or information. The degree of interpretability is largely determined by: the adequacy of definitions on concepts, target populations and variables; terminology underlying the data; and information on any limitations of the data.

Inuit

"Inuit" means "people" in Inuktitut, one of the languages of Inuit people. Most Inuit live in the Northwest Territories, Nunavut, Northern Quebec and Labrador.

Inuit Nunangat

Inuit Nunangat is the homeland of **Inuit** of Canada. It includes the communities located in the four Inuit regions: Nunatsiavut (Northern coastal Labrador), Nunavik (Northern Quebec), the territory of Nunavut and the Inuvialuit region of the Northwest Territories. These regions collectively encompass the area traditionally occupied by Inuit in Canada.

Inuk

The singular form of the word **Inuit** (i.e. 'a person').

J - K - L

Logistic regression

A form of **regression** analysis used when the response variable is a binary **variable** (a variable having two possible values).

M

Margin of error

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. In this process of **estimation**, some level of error is inevitable. The margin of error, a measure used to build confidence intervals, serves as a rough indicator of the precision of an estimate. For example, pollsters often say that a certain percentage of the population, plus or minus the margin of error (expressed in percentage points), is likely to vote for a certain candidate, 19 times out of 20. To calculate the margin of error, which in this example corresponds to a 95% confidence interval, the pollster would use the equivalent of plus or minus two standard errors of the estimate (see **Standard error**).

Methodology

A set of research methods and techniques applied to a particular field of study. At Statistics Canada, methodology refers to survey methodology.

Métis

While there is no single definition of Métis, the term generally includes any person of mixed **North American Indian** and European ancestry who self-identifies as Métis.

Microdata

Files of **records** pertaining to individual **responding** units.

N - O

National Household Survey (NHS)

This survey was introduced in 2011 as a replacement for the long census questionnaire, more widely known as Census Form 2B/2D. The NHS is designed to collect social and economic data about the Canadian population. The objective of the NHS is to provide data for small geographic areas and small population groups. For more information on the 2011 National Household Survey, please visit: http://www12.statcan.gc.ca/nhs-enm/index-eng.cfm?HPA

North American Indian

A term that describes all Aboriginal people in Canada who are not **Inuit** or **Métis**. North American Indian peoples are one of three groups of people recognized as Aboriginal in the Constitution Act, 1982. This also refers to **First Nations** people consisting of Status and non-Status **Indians**.

Observation

Data collected for a given variable about a particular **responding unit**. Examples include the specific values for a responding unit on characteristics such as age, gender or marital status—the observation might be '77', 'woman' and 'widowed'.

P

Population, target population

The complete group of units to which survey results are to apply. These units may be persons, households, businesses, institutions, etc. This is the population for which information is wanted.

Population Centre

The term population centre replaces the term urban area (as used in the Census of Population until 2006). A population centre is defined as an area with a population of at least 1,000 and no fewer than 400 persons per square kilometre. Population centres are classified into three groups, depending on the size of their population:

- small population centres, with a population between 1,000 and 29,999
- medium population centres, with a population between 30,000 and 99,999
- large urban population centres, with a population of 100,000 or more.

Postcensal survey

A postcensal survey is one where **surveyed units** are selected based upon their responses to the **Census of Population** or the **National Household Survey (NHS)**. These surveys are generally conducted shortly after the Census or NHS.

Proportion

A proportion refers to how many responses fall into a given response category in relation to the total responses.It is calculated by dividing the frequency of the response category by the total number of responses to the question.

PUMF - Public use microdata file

Public use microdata files provide access to **responding units** so that users can conduct their own research or analysis. They involve a non-identifiable data set containing characteristics pertaining to the units of the survey (e.g., individuals, households or businesses). All such datasets have been authorized for release to the public by the Statistics Canada Microdata Release Committee. The dataset contains no confidential information in that individual identifiers have been removed and any data combination or geography which could potentially reveal the identity of a responding unit has been modified.

Q – R

Record

A record is the data for an individual **responding unit** in a file containing data for all of a survey's responding units.

Regression

A statistical method which tries to predict the value of a characteristic by studying its relationship with one or more other characteristics. This relationship is expressed through the means of a regression equation.

Research Data Centres (RDCs)

The Research Data Centre Program provides researchers with access, in a secure Statistics Canada governed setting, to micro data from population and household surveys. The RDC program is part of an initiative by Statistics Canada, the Social Sciences and Humanities Research Council (SSHRC) and university consortia to help strengthen Canada's social research capacity and to support the policy research community. The program is also supported by the Canadian Foundation for Innovation (CFI) and the Canadian Institutes of Health Research (CIHR).

Respondent, responding unit

The respondent is the person providing the information for the **surveyed unit**, which could be a person, household, business or institution. In the case of the 2012 Aboriginal Peoples Survey, the respondent is the selected adult aged 18 and older or the parent or guardian of each selected child and youth aged 6 to 14 years. For youth between the ages of 15 and 17, the prior approval of the individual's parent or guardian is required in order to conduct the interview directly with the youth. Thus, for the age group 15 to 17, the respondent is either the youth or his or her parent or guardian.

Response rate

The proportion of a sample for which a response to a questionnaire is obtained, usually expressed as a percentage. Non-response covers those who refused to participate as well as persons whom the survey was unable to reach.

Sample design

A set of specifications that describe the sampling elements of a survey in detail. These elements include **population, frame, surveyed units,** sample size, sample selection and **estimation** method.

Sampling

The process of selecting some part of a population to observe so as to estimate something of interest about the whole population. Examples of different sampling methods include simple random sampling, stratified random sampling, cluster sampling, multiple-phase sampling and multi-stage sampling.

Sampling rate/Sampling fraction

Sample size divided by the population size.

Sampling or sampled unit

The unit selected by the **sample design** and from which measurements are taken for a survey. Examples include persons, households, families or businesses. For APS, the sampling unit is the person.

Standard deviation

Standard deviation measures the dispersion of a data set around the mean. It is the most widely-used measure of dispersion. Mathematically, the standard deviation is the square root of the **variance**.

Standard error

In a sample survey, results from the sample are used to estimate what the findings would be if the whole population were to be measured. **Sampling error** refers to the fact that the results of the weighted sample differ somewhat from the results that would have been obtained from the total population. The difference is known as sampling error. The actual sampling error is of course unknown, but it is possible to calculate an "average" value, known as the "standard error".

Statistics Act

An Act regarding statistics of Canada. Includes the definition of Statistics Canada's mandate: "There shall continue to be a statistics bureau under the Minister, to be known as Statistics Canada, the duties of which are:

- to collect, compile, analyze, abstract and publish statistical information relating to the commercial, industrial, financial, social, economic and general activities and condition of the people;
- to collaborate with departments of government in the collection, compilation and publication of statistical information, including statistics derived from the activities of those departments;
- to take the census of population of Canada and the census of agriculture of Canada as provided in this Act;
- to promote the avoidance of duplication in the information collected by departments of government; and,
- generally, to promote and develop integrated social and economic statistics pertaining to the whole of Canada and to each of the provinces thereof and to coordinate plans for the integration of those statistics."

Stratified sampling, stratification

A sampling procedure in which the population is divided into homogeneous subgroups or strata and the selection of samples is done independently in each stratum.

Suppression

The process by which particular data are prevented from being released based on criteria designed to protect confidentiality. 'Cell' suppression refers to procedures used to protect sensitive tabular data from disclosure; a cell being an individual entry in a table. For the APS, data were also suppressed for reasons of data quality (CV larger than 33.3%).

Surveyed unit

The selected unit from which measurements are taken for a sample survey or a Census. Examples include persons, households, families or businesses. For APS, the surveyed unit (which is also the sampled unit since the APS is a sample survey) includes children/youth aged 6 to 14 years and adults aged 15 and older.

T - U - V

User guides

These guides accompany Statistics Canada survey datasets, such as **Analytical Files** and **Public use microdata files (PUMF)**, providing the detailed technical information required to use the data appropriately. The guide typically contains important information to know prior to data analysis: weighting variables to use, procedures related to the estimate of variance, and precautions to take in the dissemination of the data.

Valid skip

Indicates that a survey question was skipped because it did not apply to the respondent's situation, as determined by valid answers to previous questions. In such cases, the respondent is not considered to be part of the target population or universe for that question.

Variable

A characteristic that may assume more than one value to which a numerical measure can be assigned (e.g. income, age and weight).

Variance

A measure of dispersion for a given characteristic or variable in a dataset. It indicates how much variability exists for that characteristic. Technically, it is calculated as the average squared deviation from the mean of each observation in the data set for a particular variable.

W - X - Y - Z

Weight

A weight is the average number of units in the population that a unit in the survey represents. Examples of a unit include a person or a household. Weights are applied to **responding units** in a sample database in order to ensure that, when making inferences from the survey data to population parameters, estimates of characteristics for the total population are obtained.

References

Bernier, J. and Nobrega, K. (1998). Outlier detection in asymmetric samples: A comparison of an inter-quartile range method and a variation of a sigma-gap method. Statistical Society of Canada Annual Meeting, June 1998.

Canadian Council on Learning (CCL). (2009). The State of Aboriginal Learning in Canada: A Holistic Approach To Measuring Success. Ottawa.

Haddou, M. (2013). Bootstrap Variance Estimation Specifications - Aboriginal Peoples Survey. Internal document, January 2013.

Langlet, É., Beaumont, J.-F., and Lavallée, P. (2008). Bootstrap Methods for Two-Phase Sampling Applicable to Postcensal Surveys. Paper presented at Statistics Canada's Advisory Committee on Statistical Methods, May 2008, Ottawa.

Verret, F. (2013). The estimation methodology of the 2011 National Household Survey.
Paper presented at the 2013 Joint Statistical Meeting.